

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:14:14 ; Search time 6 Seconds  
(without alignments)  
282.548 Million cell updates/sec

Title: US-10-057-890A-6  
Perfect score: 40  
Sequence: 1 YKGLC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	40	100.0	6	5	ABG32535	Abg32535 Scaffold
2	40	100.0	23	6	ABO13357	Abol3357 Mouse zin
3	40	100.0	25	2	AAR15297	Aar15297 Metal bin
4	40	100.0	138	5	ABG32539	Abg32539 Human CCR
5	40	100.0	157	5	ABG32540	Abg32540 Human CCR
6	40	100.0	427	6	ABU26713	Abu26713 Protein e
7	38	95.0	357	4	AAB93635	Aab93635 Human pro
8	38	95.0	357	7	ADB64416	Adb64416 Human pro
9	38	95.0	406	6	ADA54955	Ada54955 Human pro

10	38	95.0	519	6	ADA54404	Ada54404	Human pro
11	38	95.0	618	6	ADA54911	Ada54911	Human pro
12	38	95.0	620	4	AAU30585	Aau30585	Novel hum
13	38	95.0	632	4	ABG19011	Abg19011	Novel hum
14	38	95.0	769	6	AAE33782	Aae33782	Human nuc
15	37	92.5	23	6	ABO13550	Abol3550	Arabidops
16	37	92.5	139	3	AAG39378	Aag39378	Arabidops
17	37	92.5	139	3	AAG04173	Aag04173	Arabidops
18	37	92.5	149	6	ABU01311	Abu01311	S. pneumo
19	37	92.5	193	3	AAG04172	Aag04172	Arabidops
20	37	92.5	193	3	AAG39377	Aag39377	Arabidops
21	37	92.5	197	3	AAG04171	Aag04171	Arabidops
22	37	92.5	364	7	ADC37568	Adc37568	Human nuc
23	37	92.5	538	4	ABB71212	Abb71212	Drosophil
24	37	92.5	614	5	ABG95656	Abg95656	Human nuc
25	37	92.5	645	4	ABG28285	Abg28285	Novel hum
26	37	92.5	702	5	ABB04333	Abb04333	Human zin
27	37	92.5	799	3	AAB21003	Aab21003	Human nuc
28	37	92.5	799	7	ADB64775	Adb64775	Human pro
29	37	92.5	967	8	ADE52670	Ade52670	Human pro
30	36	90.0	359	3	AAB34522	Aab34522	Human sec
31	36	90.0	360	3	AAB34521	Aab34521	Human sec
32	36	90.0	360	7	ADD44831	Add44831	Rat Prote
33	36	90.0	360	7	ADD44835	Add44835	Rat Prote
34	36	90.0	360	7	ADD44839	Add44839	Rat Prote
35	36	90.0	360	7	ADD46478	Add46478	Rat Prote
36	36	90.0	360	7	ADD44843	Add44843	Rat Prote
37	36	90.0	361	3	AAB34470	Aab34470	Human sec
38	36	90.0	361	4	AAM78800	Aam78800	Human pro
39	36	90.0	361	7	ADD19004	Add19004	Human dis
40	36	90.0	362	7	ADD44833	Add44833	Human Pro
41	36	90.0	362	7	ADD44841	Add44841	Human Pro
42	36	90.0	362	7	ADD44837	Add44837	Human Pro
43	36	90.0	362	7	ADD44845	Add44845	Human Pro
44	36	90.0	362	7	ADD46480	Add46480	Human Pro
45	36	90.0	393	4	AAM79784	Aam79784	Human pro

# ALIGNMENTS

## RESULT 1

ABG32535

ID ABG32535 standard; peptide; 6 AA.

XX

AC ABG32535;

XX

DT 15-NOV-2002 (first entry)

XX

DE Scaffold domain peptide #1.

XX

KW Scaffolded protein; CCR5; HIV; human immunodeficiency virus infection;  
KW ECD; extracellular domain; metal chelating motif; zinc finger protein;  
KW integral membrane protein; soluble loop; intracellular domain; ICD;  
KW gene therapy; immunogen; viral infection; scaffold domain; human.

XX

OS Homo sapiens.

XX  
 PN WO200260477-A1.  
 XX  
 PD 08-AUG-2002.  
 XX  
 PF 29-JAN-2002; 2002WO-US002377.  
 XX  
 PR 31-JAN-2001; 2001US-0265782P.  
 PR 31-JAN-2001; 2001US-0265858P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Coleman TA, Mansfield B;  
 XX  
 DR WPI; 2002-643357/69.  
 XX  
 PT Novel scaffolded fusion polypeptide useful for therapeutic purposes or  
 PT for screening molecules that bind/activate/inhibit/modulate the  
 PT polypeptide, comprises a functional polypeptide domain fused to a  
 PT scaffold domain.  
 XX  
 PS Claim 9; Page 16; 64pp; English.  
 XX  
 CC The invention relates to a scaffolded fusion polypeptide comprising a  
 CC functional polypeptide domain fused to a scaffold domain, where the  
 CC functional polypeptide domain corresponds to a soluble loop of an  
 CC integral membrane protein (e.g. human CCR5, a transmembrane receptor  
 CC involved in HIV (human immunodeficiency virus) infection). Also included  
 CC are; (1) a polypeptide comprising a scaffold domain; (2) a nucleic acid  
 CC encoding the fusion polypeptide; (3) a vector cassette for the expression  
 CC of the fusion polypeptide comprising an expression region operably linked  
 CC to a promoter, where the expression region comprises a number of  
 CC cassettes, each of which encodes a module, domain or strand of the fusion  
 CC polypeptide and (4) a host cell comprising the vector or nucleic acid.  
 CC The fusion polypeptide is useful for screening molecules that  
 CC bind/activate/inhibit/modulate the fusion polypeptide, by expressing the  
 CC fusion polypeptide from and identifying a molecule that binds to the  
 CC fusion polypeptide. The fusion polypeptide is useful in diagnostic  
 CC methods, in assays to identify compounds that interact with loops of  
 CC fragments of an extracellular domain (ECD) or an intracellular domain  
 CC (ICD) or to rapidly assay the function of mutated portions of mutant  
 CC integral membrane proteins without having to produce significant  
 CC quantities of the entire mutant integral membrane protein, to generate  
 CC antibodies that recognise the integral membrane proteins from which they  
 CC are designed, to competitively bind the ligand of a naturally occurring  
 CC receptor in vitro or in vivo, to display and/or screen soluble domains  
 CC from protein such as integral membrane proteins, to probe the structure  
 CC of ECD or ICD, or both, of an integral protein membrane, to modulate the  
 CC activity of a receptor in vivo, and for treating or preventing viral  
 CC infection, preferably human HIV infection e.g. by gene therapy using the  
 CC encoding nucleic acid. The present sequence is a scaffold domain suitable  
 CC for inclusion in the fusion protein of the invention  
 XX  
 SQ Sequence 6 AA;

Query Match 100.0%; Score 40; DB 5; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 1 YKCGLC 6

RESULT 2

ABO13357

ID ABO13357 standard; peptide; 23 AA.

XX

AC ABO13357;

XX

DT 25-AUG-2003 (first entry)

XX

DE Mouse zinc finger DNA binding domain #463.

XX

KW Composite binding polypeptide; zinc finger nucleic acid binding domain;  
KW autoimmune disorder; immunosuppressive; zinc finger DNA binding domain;  
KW mouse.

XX

OS Mus sp.

XX

PN WO200299084-A2.

XX

PD 12-DEC-2002.

XX

PF 04-APR-2002; 2002WO-US022272.

XX

PR 04-APR-2001; 2001GB-00008491.

XX

PA (SANG-) SANGAMO BIOSCIENCES INC.

XX

PI Moore M, Sepp A, Isalan M, Choo Y;

XX

DR WPI; 2003-278214/27.

XX

PT New composite binding zinc finger polypeptide, useful for designing  
PT sequence-specific binding proteins regulating gene expression in the  
PT fields of molecular biology, and for the diagnosis and treatment of  
PT autoimmune disorders.

XX

PS Example 3; Page 107; 157pp; English.

XX

CC The invention relates to a composite binding polypeptide comprising a  
CC first natural binding domain derived from a first natural binding  
CC polypeptide and a second natural binding domain derived from a second  
CC natural binding polypeptide, where the first and second natural binding  
CC polypeptides may be the same or different and where the polypeptide binds  
CC to a target differing from the natural target of both the first and  
CC second binding polypeptides. The invention also relates to a chimeric  
CC polypeptide comprising a binding polypeptide cited above and a biological  
CC effector domain, a library of natural binding domains, a library of  
CC natural zinc finger nucleic acid binding domains comprising a linker  
CC attached to it, a method for selecting a binding polypeptide capable of  
CC binding to a target site and a method for designing a composite binding  
CC polypeptide. The methods and compositions of the present invention are



CC useful for designing sequence-specific binding proteins for regulation of  
CC gene expression in the fields of molecular biology. They can also be used  
CC for the diagnosis and treatment of autoimmune disorders, and as research  
CC tools and in transgenic animals. This sequence represents a mouse zinc  
CC finger DNA binding domain used in the scope of the invention  
XX  
SQ Sequence 23 AA;

Query Match 100.0%; Score 40; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 5.4;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| | | | |  
Db 1 YKCGLC 6

### RESULT 3

AAR15297

ID AAR15297 standard; protein; 25 AA.

XX

AC AAR15297;

XX

DT 17-FEB-1992 (first entry)

XX

DE Metal binding effector domain (6).

XX

KW Molecular recognition unit; MRU; platelet fibrinogen receptor; iron;  
KW effector domain.

XX

OS Synthetic.

XX

PN WO9117173-A.

XX

PD 14-NOV-1991.

XX

PF 07-MAY-1990; 90US-00519702.

XX

PR 07-MAY-1990; 90US-00519702.

XX

PA (CYTO-) CYTOGEN CORP.

XX

PI Rodwell JD, Mckearn TJ, Alvarez VL, Radcliffe RD;

XX

DR WPI; 1991-353714/48.

XX

PT New conjugates of molecular recognition units - having first binding  
PT domain for desired target site and second effector domain, useful in  
PT medicine, agriculture, waste management etc.

XX

PS Claim 8; Page 60; 74pp; English.

XX

CC A conjugate of an MRU comprises a fusion peptide having a first binding  
CC domain (e.g. having the sequence represented in AAR15289-91) which is an  
CC MRU having binding specificity and affinity for an activated platelet  
CC fibrinogen receptor and a second effector domain (e.g. having the  
CC sequence represented in AAR15292-97) which has the ability to bind iron.

CC The amino or carboxy terminus of the peptide can be blocked. See also  
CC AAR15373-97  
XX  
SQ Sequence 25 AA;

Query Match 100.0%; Score 40; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| | | | |  
Db 1 YKCGLC 6

RESULT 4

ABG32539

ID ABG32539 standard; protein; 138 AA.

XX

AC ABG32539;

XX

DT 15-NOV-2002 (first entry)

XX

DE Human CCR5-based scaffolded fusion protein #1.

XX

KW Scaffolded protein; CCR5; HIV; human immunodeficiency virus infection;

KW ECD; extracellular domain; metal chelating motif; zinc finger protein;

KW integral membrane protein; soluble loop; intracellular domain; ICD;

KW gene therapy; immunogen; viral infection; human.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200260477-A1.

XX

PD 08-AUG-2002.

XX

PF 29-JAN-2002; 2002WO-US002377.

XX

PR 31-JAN-2001; 2001US-0265782P.

PR 31-JAN-2001; 2001US-0265858P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Coleman TA, Mansfield B;

XX

DR WPI; 2002-643357/69.

XX

PT Novel scaffolded fusion polypeptide useful for therapeutic purposes or

PT for screening molecules that bind/activate/inhibit/modulate the

PT polypeptide, comprises a functional polypeptide domain fused to a

PT scaffold domain.

XX

PS Example 1; Page 21; 64pp; English.

XX

CC The invention relates to a scaffolded fusion polypeptide comprising a

CC functional polypeptide domain fused to a scaffold domain, where the

CC functional polypeptide domain corresponds to a soluble loop of an

CC integral membrane protein (e.g. human CCR5, a transmembrane receptor  
 CC involved in HIV (human immunodeficiency virus) infection). Also included  
 CC are; (1) a polypeptide comprising a scaffold domain; (2) a nucleic acid  
 CC encoding the fusion polypeptide; (3) a vector cassette for the expression  
 CC of the fusion polypeptide comprising an expression region operably linked  
 CC to a promoter, where the expression region comprises a number of  
 CC cassettes, each of which encodes a module, domain or strand of the fusion  
 CC polypeptide and (4) a host cell comprising the vector or nucleic acid.  
 CC The fusion polypeptide is useful for screening molecules that  
 CC bind/activate/inhibit/modulate the fusion polypeptide, by expressing the  
 CC fusion polypeptide from and identifying a molecule that binds to the  
 CC fusion polypeptide. The fusion polypeptide is useful in diagnostic  
 CC methods, in assays to identify compounds that interact with loops of  
 CC fragments of an extracellular domain (ECD) or an intracellular domain  
 CC (ICD) or to rapidly assay the function of mutated portions of mutant  
 CC integral membrane proteins without having to produce significant  
 CC quantities of the entire mutant integral membrane protein, to generate  
 CC antibodies that recognise the integral membrane proteins from which they  
 CC are designed, to competitively bind the ligand of a naturally occurring  
 CC receptor in vitro or in vivo, to display and/or screen soluble domains  
 CC from protein such as integral membrane proteins, to probe the structure  
 CC of ECD or ICD, or both, of an integral protein membrane, to modulate the  
 CC activity of a receptor in vivo, and for treating or preventing viral  
 CC infection, preferably human HIV infection e.g. by gene therapy using the  
 CC encoding nucleic acid. The present sequence is a scaffolded protein based  
 CC on the ECD region of human CCR5 (not defined)

XX

SQ Sequence 138 AA;

Query Match 100.0%; Score 40; DB 5; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||  
 Db 31 YKCGLC 36

# RESULT 5

ABG32540

ID ABG32540 standard; protein; 157 AA.

XX

AC ABG32540;

XX

DT 15-NOV-2002 (first entry)

XX

DE Human CCR5-based scaffolded fusion protein #2.

XX

KW Scaffolded protein; CCR5; HIV; human immunodeficiency virus infection;  
 KW ECD; extracellular domain; metal chelating motif; zinc finger protein;  
 KW integral membrane protein; soluble loop; intracellular domain; ICD;  
 KW gene therapy; immunogen; viral infection; human.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 1. .19  
FT /label= Signal\_peptide  
FT Protein 20. .157  
FT /label= Mature\_scaffolded\_protein

XX

PN WO200260477-A1.

XX

PD 08-AUG-2002.

XX

PF 29-JAN-2002; 2002WO-US002377.

XX

PR 31-JAN-2001; 2001US-0265782P.

PR 31-JAN-2001; 2001US-0265858P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Coleman TA, Mansfield B;

XX

DR WPI; 2002-643357/69.

XX

PT Novel scaffolded fusion polypeptide useful for therapeutic purposes or  
PT for screening molecules that bind/activate/inhibit/modulate the  
PT polypeptide, comprises a functional polypeptide domain fused to a  
PT scaffold domain.

XX

PS Example 2; Page 41; 64pp; English.

XX

CC The invention relates to a scaffolded fusion polypeptide comprising a  
CC functional polypeptide domain fused to a scaffold domain, where the  
CC functional polypeptide domain corresponds to a soluble loop of an  
CC integral membrane protein (e.g. human CCR5, a transmembrane receptor  
CC involved in HIV (human immunodeficiency virus) infection). Also included  
CC are; (1) a polypeptide comprising a scaffold domain; (2) a nucleic acid  
CC encoding the fusion polypeptide; (3) a vector cassette for the expression  
CC of the fusion polypeptide comprising an expression region operably linked  
CC to a promoter, where the expression region comprises a number of  
CC cassettes, each of which encodes a module, domain or strand of the fusion  
CC polypeptide and (4) a host cell comprising the vector or nucleic acid.  
CC The fusion polypeptide is useful for screening molecules that  
CC bind/activate/inhibit/modulate the fusion polypeptide, by expressing the  
CC fusion polypeptide from and identifying a molecule that binds to the  
CC fusion polypeptide. The fusion polypeptide is useful in diagnostic  
CC methods, in assays to identify compounds that interact with loops of  
CC fragments of an extracellular domain (ECD) or an intracellular domain  
CC (ICD) or to rapidly assay the function of mutated portions of mutant  
CC integral membrane proteins without having to produce significant  
CC quantities of the entire mutant integral membrane protein, to generate  
CC antibodies that recognise the integral membrane proteins from which they  
CC are designed, to competitively bind the ligand of a naturally occurring  
CC receptor in vitro or in vivo, to display and/or screen soluble domains  
CC from protein such as integral membrane proteins, to probe the structure  
CC of ECD or ICD, or both, of an integral protein membrane, to modulate the  
CC activity of a receptor in vivo, and for treating or preventing viral  
CC infection, preferably human HIV infection e.g. by gene therapy using the  
CC encoding nucleic acid. The present sequence is a scaffolded protein based  
CC on the ECD region of human CCR5 (not defined)

XX

SQ Sequence 157 AA;

Query Match 100.0%; Score 40; DB 5; Length 157;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
| | | | |  
Db 50 YKGLC 55

RESULT 6

ABU26713

ID ABU26713 standard; protein; 427 AA.

XX

AC ABU26713;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #12240.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Campylobacter jejuni.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

DR N-PSDB; ACA30583.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 25; SEQ ID NO 54637; 1766pp; English.

XX

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 427 AA;

Query Match 100.0%; Score 40; DB 6; Length 427;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||  
 Db 86 YKCGLC 91

# RESULT 7

AAB93635

ID AAB93635 standard; protein; 357 AA.

XX

AC AAB93635;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:13123.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.

XX

PS Claim 8; SEQ ID NO 13123; 2537pp + Sequence Listing; English.

XX

CC The present invention describes primer sets for synthesising 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

XX

SQ Sequence 357 AA;

Query Match 95.0%; Score 38; DB 4; Length 357;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKCGLC 6  
| | | | : |  
Db 334 YKCGIC 339

RESULT 8

ADB64416

ID ADB64416 standard; protein; 357 AA.

XX  
 AC ADB64416;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human protein encoded by clone FEBRA20027070.  
 XX  
 KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 XX  
 DR WPI; 2003-450961/43.  
 DR N-PSDB; ADB62446.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX  
 PS Claim 1; Page; 222pp; English.  
 XX  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,



CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a protein of the invention. Note: Some of the  
CC sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

XX

SQ Sequence 357 AA;

Query Match 95.0%; Score 38; DB 7; Length 357;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| | | | : |  
Db 334 YKCGIC 339

RESULT 9

ADA54955

ID ADA54955 standard; protein; 406 AA.

XX

AC ADA54955;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human protein, SEQ ID 2523.

XX

KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

KW inflammatory disease; osteoporosis; neurological disease.

XX

OS Homo sapiens.

XX

PN EP1293569-A2.

XX

PD 19-MAR-2003.

XX

PF 21-MAR-2002; 2002EP-00006586.

XX

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

XX

PA (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX

DR WPI; 2003-395539/38.

DR N-PSDB; ADA53316.

XX

PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.

XX

PS Claim 14; SEQ ID NO 2523; 205pp; English.

XX

CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.

XX

SQ Sequence 406 AA;

Query Match 95.0%; Score 38; DB 6; Length 406;

Best Local Similarity 83.3%; Pred. No. 1.5e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

||||:

Db 383 YKCGIC 388

#### RESULT 10

ADA54404

ID ADA54404 standard; protein; 519 AA.

XX

AC ADA54404;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human protein, SEQ ID 1972.

XX

KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

KW inflammatory disease; osteoporosis; neurological disease.

XX

OS Homo sapiens.

XX

PN EP1293569-A2.

XX

PD 19-MAR-2003.

XX

PF 21-MAR-2002; 2002EP-00006586.

XX

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

XX

PA (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX

DR WPI; 2003-395539/38.

DR N-PSDB; ADA52765.

XX

PT New polynucleotides encoding full-length polypeptides, e.g. secretory

PT and/or membrane proteins, useful for developing medicines for diseases in

PT which the gene is involved, or as target molecules for gene therapy.

XX  
PS Claim 14; SEQ ID NO 1972; 205pp; English.

XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.

XX  
SQ Sequence 519 AA;

Query Match 95.0%; Score 38; DB 6; Length 519;  
Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| | | | : |  
Db 435 YKCGMC 440

RESULT 11

ADA54911

ID ADA54911 standard; protein; 618 AA.

XX

AC ADA54911;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human protein, SEQ ID 2479.

XX

KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

KW inflammatory disease; osteoporosis; neurological disease.

XX

OS Homo sapiens.

XX

PN EP1293569-A2.

XX

PD 19-MAR-2003.

XX

PF 21-MAR-2002; 2002EP-00006586.

XX

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

XX

PA (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX

DR WPI; 2003-395539/38.

DR N-PSDB; ADA53272.

XX

PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in

PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 14; SEQ ID NO 2479; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 618 AA;

Query Match 95.0%; Score 38; DB 6; Length 618;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| | | | : |  
Db 220 YKCGIC 225

RESULT 12

AAU30585

ID AAU30585 standard; protein; 620 AA.

XX

AC AAU30585;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #1076.

XX

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200179449-A2.

XX

PD 25-OCT-2001.

XX

PF 16-APR-2001; 2001WO-US008656.

XX

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-611725/70.

XX

PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
PT vaccination, testing and therapy.

XX

PS Claim 20; Page 314; 765pp; English.

XX

CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising the  
 CC nucleic acids encoding the polypeptides and cells genetically engineered  
 CC to express them are also useful for producing the proteins. The proteins  
 CC are useful in genetic vaccination, testing and therapy, and can be used  
 CC as nutritional supplements. They may be used to increase stem cell  
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
 CC and/or nerve tissue growth or regeneration; immune suppression and/or  
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human  
 CC secreted proteins of the invention  
 XX  
 SQ Sequence 620 AA;

Query Match 95.0%; Score 38; DB 4; Length 620;  
 Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 ||||:|  
 Db 359 YKCGIC 364

# RESULT 13

ABG19011

ID ABG19011 standard; protein; 632 AA.

XX

AC ABG19011;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #19002.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS83198.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX

PS Claim 20; SEQ ID NO 49370; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 632 AA;

Query Match 95.0%; Score 38; DB 4; Length 632;

Best Local Similarity 83.3%; Pred. No. 2.2e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKCGLC 6

||||:

Db 609 YKCGIC 614

RESULT 14

AAE33782

ID AAE33782 standard; protein; 769 AA.

XX

AC AAE33782;

XX

DT 16-APR-2003 (first entry)

XX

DE Human nucleic acid associated protein (NAAP)-22.

XX

KW Human; nucleic acid associated protein; NAAP; cancer; atherosclerosis;  
KW cell proliferative disorder; neurological disorder; Huntington's disease;  
KW epilepsy; stroke; immune disorder; acquired immune deficiency syndrome;  
KW AIDS; inflammatory disorder; allergy; developmental disorder; infection;

KW   hypothyroidism; cushing's syndrome; gene therapy; cytostatic; nootropic;  
KW   anticonvulsant; neuroprotective; cerebroprotective; thyromimetic.  
XX  
OS   Homo sapiens.  
XX  
FH   Key                    Location/Qualifiers  
FT   Domain                24. .85  
FT                         /note= "Krab box domain"  
FT   Domain                26. .97  
FT                         /note= "Krab box domain"  
FT   Domain                160. .180  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                188. .208  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                216. .236  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                244. .264  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                272. .292  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                300. .320  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                356. .376  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                384. .404  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                412. .432  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                413. .428  
FT                         /note= "Transmembrane domain"  
FT   Domain                440. .460  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                468. .488  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                496. .516  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                524. .544  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                552. .572  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                580. .600  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                608. .628  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                636. .656  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                664. .684  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                692. .712  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                720. .740  
FT                         /note= "Zinc finger C2H2 type domain"  
FT   Domain                748. .768  
FT                         /note= "Zinc finger C2H2 type domain"  
XX  
PN   WO200299115-A2.  
XX

PD 12-DEC-2002.  
 XX  
 PF 31-MAY-2002; 2002WO-US017050.  
 XX  
 PR 01-JUN-2001; 2001US-0295359P.  
 PR 08-JUN-2001; 2001US-0296878P.  
 PR 08-JUN-2001; 2001US-0297222P.  
 PR 15-JUN-2001; 2001US-0298615P.  
 PR 15-JUN-2001; 2001US-0298665P.  
 PR 15-JUN-2001; 2001US-0298693P.  
 PR 21-JUN-2001; 2001US-0300176P.  
 PR 19-APR-2002; 2002US-0373891P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PA (YUEH/) YUE H.  
 XX  
 PI Yue H, Tang YT, Baughn MR, Becha SD, Warren BA, Walia NK;  
 PI Lal PG, Lee EA, Hafalia AJA, Richardson TW, Griffin JA, Emerling BM;  
 PI Ramkumar J, Yue H, Swarnakar A, Tran B, Li JX, Yao MG, Yang J;  
 PI Ison CH, Forsythe IJ, Honchell CD, Arvizu CS, Elliott VS, Lu Y;  
 PI Ding L, Luo W, Wang YE, Burford N, Borowsky ML, Nguyen DB;  
 PI Chinn AM, Kable AE;  
 XX  
 DR WPI; 2003-140626/13.  
 DR N-PSDB; AAD51696.  
 XX  
 PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
 PT treating and preventing diseases or conditions associated with the  
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
 PT infections.  
 XX  
 PS Claim 78; Col 214-215; 257pp; English.  
 XX  
 CC The present invention relates to human nucleic acid associated proteins  
 CC (NAAP) and polynucleotides encoding such proteins. NAAP sequences are  
 CC useful in diagnosing, treating and preventing diseases or conditions  
 CC associated with the decreased expression or overexpression of NAAP such  
 CC as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
 CC disorders or infections. They are also used in gene therapy. The present  
 CC sequence is human NAAP-22 protein  
 XX  
 SQ Sequence 769 AA;  
  
 Query Match 95.0%; Score 38; DB 6; Length 769;  
 Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 Qy 1 YKCGLC 6  
 ||||:|  
 Db 746 YKCGIC 751  
  
 RESULT 15  
 ABO13550  
 ID ABO13550 standard; peptide; 23 AA.



XX  
 AC AB013550;  
 XX  
 DT 25-AUG-2003 (first entry)  
 XX  
 DE Arabidopsis zinc finger DNA binding domain #23.  
 XX  
 KW Composite binding polypeptide; zinc finger nucleic acid binding domain;  
 KW autoimmune disorder; immunosuppressive; zinc finger DNA binding domain.  
 XX  
 OS Arabidopsis sp.  
 XX  
 PN WO200299084-A2.  
 XX  
 PD 12-DEC-2002.  
 XX  
 PF 04-APR-2002; 2002WO-US022272.  
 XX  
 PR 04-APR-2001; 2001GB-00008491.  
 XX  
 PA (SANG-) SANGAMO BIOSCIENCES INC.  
 XX  
 PI Moore M, Sepp A, Isalan M, Choo Y;  
 XX  
 DR WPI; 2003-278214/27.  
 XX  
 PT New composite binding zinc finger polypeptide, useful for designing  
 PT sequence-specific binding proteins regulating gene expression in the  
 PT fields of molecular biology, and for the diagnosis and treatment of  
 PT autoimmune disorders.  
 XX  
 PS Example 3; Page 111; 157pp; English.  
 XX  
 CC The invention relates to a composite binding polypeptide comprising a  
 CC first natural binding domain derived from a first natural binding  
 CC polypeptide and a second natural binding domain derived from a second  
 CC natural binding polypeptide, where the first and second natural binding  
 CC polypeptides may be the same or different and where the polypeptide binds  
 CC to a target differing from the natural target of both the first and  
 CC second binding polypeptides. The invention also relates to a chimeric  
 CC polypeptide comprising a binding polypeptide cited above and a biological  
 CC effector domain, a library of natural binding domains, a library of  
 CC natural zinc finger nucleic acid binding domains comprising a linker  
 CC attached to it, a method for selecting a binding polypeptide capable of  
 CC binding to a target site and a method for designing a composite binding  
 CC polypeptide. The methods and compositions of the present invention are  
 CC useful for designing sequence-specific binding proteins for regulation of  
 CC gene expression in the fields of molecular biology. They can also be used  
 CC for the diagnosis and treatment of autoimmune disorders, and as research  
 CC tools and in transgenic animals. This sequence represents an Arabidopsis  
 CC zinc finger DNA binding domain used in the scope of the invention  
 XX  
 SQ Sequence 23 AA;

Query Match 92.5%; Score 37; DB 6; Length 23;  
 Best Local Similarity 83.3%; Pred. No. 17;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy            1 YKGLC 6  
              ||||:|  
Db            1 YKGV 6

Search completed: March 5, 2004, 16:22:44  
Job time : 7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:17:14 ; Search time 1.61111 Seconds  
(without alignments)  
192.262 Million cell updates/sec

Title: US-10-057-890A-6  
Perfect score: 40  
Sequence: 1 YKGLC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	40	100.0	25	3	US-09-253-396A-216	Sequence 216, App
2	40	100.0	25	6	5196510-11	Patent No. 5196510
3	40	100.0	26	2	US-08-620-151-55	Sequence 55, Appl
4	35	87.5	127	4	US-09-107-532A-4659	Sequence 4659, Ap
5	35	87.5	313	4	US-09-800-729-196	Sequence 196, App
6	34	85.0	27	2	US-08-620-151-76	Sequence 76, Appl
7	34	85.0	29	3	US-09-253-396A-217	Sequence 217, App
8	34	85.0	61	4	US-09-277-431A-4	Sequence 4, Appli
9	34	85.0	61	4	US-09-277-431A-5	Sequence 5, Appli
10	34	85.0	81	4	US-09-543-681A-6176	Sequence 6176, Ap
11	34	85.0	310	4	US-09-252-991A-27339	Sequence 27339, A

12	33	82.5	70	4	US-09-621-976-5975	Sequence 5975, Ap
13	33	82.5	352	4	US-08-979-847B-118	Sequence 118, App
14	33	82.5	378	4	US-08-979-847B-122	Sequence 122, App
15	33	82.5	398	4	US-08-979-847B-121	Sequence 121, App
16	33	82.5	399	4	US-09-252-991A-18242	Sequence 18242, A
17	33	82.5	488	2	US-08-933-750C-17	Sequence 17, Appl
18	33	82.5	488	3	US-09-234-613-17	Sequence 17, Appl
19	33	82.5	577	2	US-07-728-215-29	Sequence 29, Appl
20	33	82.5	577	4	US-08-938-085A-29	Sequence 29, Appl
21	33	82.5	577	4	US-10-072-844-29	Sequence 29, Appl
22	33	82.5	577	4	US-10-072-838-29	Sequence 29, Appl
23	33	82.5	577	4	US-10-072-841A-29	Sequence 29, Appl
24	33	82.5	799	1	US-08-054-077C-2	Sequence 2, Appli
25	32	80.0	26	2	US-08-620-151-38	Sequence 38, Appl
26	32	80.0	28	2	US-08-398-590A-45	Sequence 45, Appl
27	32	80.0	28	3	US-08-894-997-45	Sequence 45, Appl
28	32	80.0	29	2	US-08-616-857-4	Sequence 4, Appli
29	32	80.0	95	4	US-09-029-755C-7	Sequence 7, Appli
30	32	80.0	224	2	US-08-616-857-2	Sequence 2, Appli
31	32	80.0	224	4	US-09-546-013-22	Sequence 22, Appl
32	32	80.0	304	4	US-09-489-039A-12220	Sequence 12220, A
33	32	80.0	313	4	US-09-543-681A-7456	Sequence 7456, Ap
34	32	80.0	320	4	US-09-489-039A-9398	Sequence 9398, Ap
35	32	80.0	338	4	US-09-543-681A-6123	Sequence 6123, Ap
36	32	80.0	402	4	US-09-029-755C-5	Sequence 5, Appli
37	32	80.0	409	4	US-09-029-755C-2	Sequence 2, Appli
38	32	80.0	676	2	US-08-398-590A-40	Sequence 40, Appl
39	32	80.0	676	3	US-08-894-997-40	Sequence 40, Appl
40	32	80.0	976	3	US-08-894-997-50	Sequence 50, Appl
41	32	80.0	1332	2	US-08-971-244-2	Sequence 2, Appli
42	32	80.0	1332	3	US-09-286-891-2	Sequence 2, Appli
43	31	77.5	26	2	US-08-620-151-53	Sequence 53, Appl
44	31	77.5	41	4	US-09-389-831-15	Sequence 15, Appl
45	31	77.5	43	1	US-08-208-108-7	Sequence 7, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-253-396A-216

; Sequence 216, Application US/09253396A

; Patent No. 6205404

; GENERAL INFORMATION:

; APPLICANT: Genome Dynamics, Inc.

; TITLE OF INVENTION: DNA-Binding Proteins of the Zinc-Finger Class

; FILE REFERENCE: 1116242-0003 file: genome03F.app

; CURRENT APPLICATION NUMBER: US/09/253,396A

; CURRENT FILING DATE: 1999-02-19

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 216

; LENGTH: 25

; TYPE: PRT

; ORGANISM: Xenopus laevis

US-09-253-396A-216

Query Match 100.0%; Score 40; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.99;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 1 YKCGLC 6

RESULT 2

5196510-11

; Patent No. 5196510

; APPLICANT: RODWELL, JOHN D.; MCKEARN, THOMAS J.; ALVAREZ,  
; VERNON, L.; RADCLIFFE, ROBERT D.

; TITLE OF INVENTION: MOLECULAR RECOGNITION UNITS

; NUMBER OF SEQUENCES: 43

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/519,702

; FILING DATE: 07-MAY-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 291,730

; FILING DATE: 29-DEC-1988

; SEQ ID NO: 11:

; LENGTH: 25

5196510-11

Query Match 100.0%; Score 40; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.99;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 1 YKCGLC 6

RESULT 3

US-08-620-151-55

; Sequence 55, Application US/08620151

; Patent No. 5928955

; GENERAL INFORMATION:

; APPLICANT: Imperiali, Barbara

; APPLICANT: Walkup, Grant K.

; TITLE OF INVENTION: PEPTIDYL FLUORESCENT CHEMOSENSOR FOR

; TITLE OF INVENTION: DIVALENT ZINC

; NUMBER OF SEQUENCES: 136

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BRINKS, HOFER, GILSON & LIONE

; STREET: NBC Tower - Suite 3600, 455 N. Cityfront

; STREET: Plaza Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60611-5599

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/620,151
; FILING DATE: 22-MAR-1996
; CLASSIFICATION: 422
; ATTORNEY/AGENT INFORMATION:
; NAME: Shannon, Karen L.
; REGISTRATION NUMBER: 36,675
; REFERENCE/DOCKET NUMBER: 8597/6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-321-4200
; TELEFAX: 312-321-4299
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-620-151-55

```

```

Query Match          100.0%; Score 40; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 1;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 YKCGLC 6
        |||||
Db      2 YKCGLC 7

```

#### RESULT 4

US-09-107-532A-4659

; Sequence 4659, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND

#### THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

```

;          FILING DATE: 14 May 1998
;          APPLICATION NUMBER: 60/051571
;          FILING DATE: July 2, 1997
;          ATTORNEY/AGENT INFORMATION:
;          NAME: Ariniello, Pamela Deneke
;          REGISTRATION NUMBER: 40,489
;          REFERENCE/DOCKET NUMBER: GTC-012
;          TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (781)893-5007
;          TELEFAX: (781)893-8277
;          INFORMATION FOR SEQ ID NO: 4659:
;          SEQUENCE CHARACTERISTICS:
;          LENGTH: 127 amino acids
;          TYPE: amino acid
;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          HYPOTHETICAL: YES
;          ORIGINAL SOURCE:
;          ORGANISM: Enterococcus faecium
;          FEATURE:
;          NAME/KEY: misc_feature
;          LOCATION: (B) LOCATION 1...127
;          SEQUENCE DESCRIPTION: SEQ ID NO: 4659:
US-09-107-532A-4659

```

```

Query Match          87.5%;  Score 35;  DB 4;  Length 127;
Best Local Similarity 66.7%;  Pred. No. 31;
Matches      4;  Conservative      2;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

Qy      1 YKCGLC 6
        |:|:|
Db      7 YRCGIC 12

```

```

RESULT 5
US-09-800-729-196
; Sequence 196, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: PZ044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 196
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-196

```

```

Query Match          87.5%;  Score 35;  DB 4;  Length 313;

```

Best Local Similarity 83.3%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 208 YKCGAC 213

RESULT 6

US-08-620-151-76

; Sequence 76, Application US/08620151

; Patent No. 5928955

; GENERAL INFORMATION:

; APPLICANT: Imperiali, Barbara

; APPLICANT: Walkup, Grant K.

; TITLE OF INVENTION: PEPTIDYL FLUORESCENT CHEMOSENSOR FOR

; TITLE OF INVENTION: DIVALENT ZINC

; NUMBER OF SEQUENCES: 136

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BRINKS, HOFER, GILSON & LIONE

; STREET: NBC Tower - Suite 3600, 455 N. Cityfront

; STREET: Plaza Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60611-5599

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/620,151

; FILING DATE: 22-MAR-1996

; CLASSIFICATION: 422

; ATTORNEY/AGENT INFORMATION:

; NAME: Shannon, Karen L.

; REGISTRATION NUMBER: 36,675

; REFERENCE/DOCKET NUMBER: 8597/6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-321-4200

; TELEFAX: 312-321-4299

; INFORMATION FOR SEQ ID NO: 76:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 27 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-620-151-76

Query Match 85.0%; Score 34; DB 2; Length 27;

Best Local Similarity 83.3%; Pred. No. 11;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| ||||



Db

2 YPCGLC 7

RESULT 7

US-09-253-396A-217

; Sequence 217, Application US/09253396A

; Patent No. 6205404

; GENERAL INFORMATION:

; APPLICANT: Genome Dynamics, Inc.

; TITLE OF INVENTION: DNA-Binding Proteins of the Zinc-Finger Class

; FILE REFERENCE: 1116242-0003 file: genome03F.app

; CURRENT APPLICATION NUMBER: US/09/253,396A

; CURRENT FILING DATE: 1999-02-19

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 217

; LENGTH: 29

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-253-396A-217

Query Match 85.0%; Score 34; DB 3; Length 29;

Best Local Similarity 83.3%; Pred. No. 12;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

| | | | |

Db 6 YPCGLC 11

RESULT 8

US-09-277-431A-4

; Sequence 4, Application US/09277431A

; Patent No. 6656705

; GENERAL INFORMATION:

; APPLICANT: Baden, Howard P.

; APPLICANT: Olson, Pamela

; APPLICANT: Champliand, Marie-France

; TITLE OF INVENTION: SCIELLIN AND USES THEREOF

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/277,431A

; FILING DATE: 26-MAR-1999

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/079,498

```
; FILING DATE: 26-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10284/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-277-431A-4
```

```
Query Match          85.0%; Score 34; DB 4; Length 61;
Best Local Similarity 66.7%; Pred. No. 23;
Matches      4; Conservative 2; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy      1 YKCGLC 6
        :|||:|
Db      29 FKCGIC 34
```

RESULT 9

```
US-09-277-431A-5
; Sequence 5, Application US/09277431A
; Patent No. 6656705
; GENERAL INFORMATION:
; APPLICANT: Baden, Howard P.
; APPLICANT: Olson, Pamela
; APPLICANT: Champliand, Marie-France
; TITLE OF INVENTION: SCIELLIN AND USES THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/277,431A
; FILING DATE: 26-MAR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/079,498
; FILING DATE: 26-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
```

; REFERENCE/DOCKET NUMBER: 10284/009001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 61 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-277-431A-5

Query Match 85.0%; Score 34; DB 4; Length 61;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
:|||||  
Db 29 FKGLC 34

RESULT 10

US-09-543-681A-6176  
; Sequence 6176, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS  
MIRABILIS FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 6176  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-6176

Query Match 85.0%; Score 34; DB 4; Length 81;  
Best Local Similarity 83.3%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
| |||||  
Db 66 YHGLC 71

RESULT 11

US-09-252-991A-27339  
; Sequence 27339, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

```
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27339
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27339
```

```
Query Match          85.0%; Score 34; DB 4; Length 310;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches      5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 YKCGLC 6
        ||| ||
Db      155 YKCSLC 160
```

#### RESULT 12

```
US-09-621-976-5975
; Sequence 5975, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5975
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-5975
```

```
Query Match          82.5%; Score 33; DB 4; Length 70;
Best Local Similarity 100.0%; Pred. No. 39;
Matches      5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 KCGLC 6
        |||||
Db      62 KCGLC 66
```

#### RESULT 13

US-08-979-847B-118  
 ; Sequence 118, Application US/08979847B  
 ; Patent No. 6582703  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PERRON, HERVE  
 ; BESEME, FREDERIC  
 ; BEDIN, FREDERIC  
 ; PARANHOS-BACCALA, GLAUCIA  
 ; KOMURIAN-PRADEL, FLORENCE  
 ; JOLIVET-REYNAUD, COLETTE  
 ; MANDRAND, BERNARD  
 ; GARSON, JEREMY  
 ; TUKE, PHILIP  
 ; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
 ; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC,  
 PROPHYLACTIC AND  
 ; THERAPEUTIC PURPOSES  
 ; NUMBER OF SEQUENCES: 210  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OLIFF & BERRIDGE, PLC  
 ; STREET: P.O. BOX 19928  
 ; CITY: ALEXANDRIA  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22320  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/979,847B  
 ; FILING DATE: 26-No. 6582703-1997  
 ; CLASSIFICATION: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BERRIDGE, WILLIAM P.  
 ; REGISTRATION NUMBER: 30,024  
 ; REFERENCE/DOCKET NUMBER: WPB 39046A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-836-6400  
 ; TELEFAX: 703-836-2787  
 ; INFORMATION FOR SEQ ID NO: 118:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 352 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 118:  
 US-08-979-847B-118

Query Match 82.5%; Score 33; DB 4; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KCGLC 6  
 |||||

## RESULT 14

US-08-979-847B-122

; Sequence 122, Application US/08979847B

; Patent No. 6582703

## ; GENERAL INFORMATION:

; APPLICANT: PERRON, HERVE

; BESEME, FREDERIC

; BEDIN, FREDERIC

; PARANHOS-BACCALA, GLAUCIA

; KOMURIAN-PRADEL, FLORENCE

; JOLIVET-REYNAUD, COLETTE

; MANDRAND, BERNARD

; GARSON, JEREMY

; TUKE, PHILIP

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC,  
PROPHYLACTIC AND

; THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 210

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF &amp; BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: ALEXANDRIA

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/979,847B

; FILING DATE: 26-No. 6582703-1997

; CLASSIFICATION: &lt;Unknown&gt;

; ATTORNEY/AGENT INFORMATION:

; NAME: BERRIDGE, WILLIAM P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 39046A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 122:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 378 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 122:

US-08-979-847B-122

Query Match 82.5%; Score 33; DB 4; Length 378;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 KCGLC 6  
|||||  
Db 98 KCGLC 102

RESULT 15

US-08-979-847B-121

; Sequence 121, Application US/08979847B

; Patent No. 6582703

; GENERAL INFORMATION:

; APPLICANT: PERRON, HERVE

; BESEME, FREDERIC

; BEDIN, FREDERIC

; PARANHOS-BACCALA, GLAUCIA

; KOMURIAN-PRADEL, FLORENCE

; JOLIVET-REYNAUD, COLETTE

; MANDRAND, BERNARD

; GARSON, JEREMY

; TUKE, PHILIP

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC,

PROPHYLACTIC AND

; THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 210

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: ALEXANDRIA

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/979,847B

; FILING DATE: 26-No. 6582703-1997

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: BERRIDGE, WILLIAM P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 39046A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 121:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 398 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 121:

US-08-979-847B-121

Query Match 82.5%; Score 33; DB 4; Length 398;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KCGLC 6  
|||||  
Db 118 KCGLC 122

Search completed: March 5, 2004, 16:30:35  
Job time : 2.61111 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:16:19 ; Search time 1.37037 Seconds  
(without alignments)  
421.163 Million cell updates/sec

Title: US-10-057-890A-6  
Perfect score: 40  
Sequence: 1 YKCGLC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	40	100.0	427	2	G81260	UDP-N-acetylmuramo
2	40	100.0	614	2	JH0500	zinc finger protei
3	40	100.0	1350	2	S00647	finger protein - A
4	37	92.5	57	2	B43284	zinc finger protei
5	37	92.5	149	2	A95105	conserved hypothet
6	37	92.5	149	2	A97973	conserved hypothet
7	36	90.0	1638	2	T25352	hypothetical prote
8	35	87.5	582	2	C90165	conserved hypothet
9	35	87.5	1000	2	S74069	hypothetical prote
10	35	87.5	1115	2	T11614	probable poly(A)-s
11	34	85.0	186	2	T22278	hypothetical prote
12	34	85.0	190	2	T22279	hypothetical prote
13	34	85.0	277	2	T27824	hypothetical prote

14	34	85.0	309	2	F83044	nitrate-inducible
15	34	85.0	396	2	C84797	hypothetical prote
16	34	85.0	472	2	T34222	transcription fact
17	34	85.0	506	2	C64377	hypothetical prote
18	34	85.0	610	2	G72494	hypothetical prote
19	34	85.0	621	2	T22904	hypothetical prote
20	34	85.0	1028	2	A56038	DNA-binding protei
21	34	85.0	1186	2	T33754	O/E-1-associated z
22	34	85.0	1213	2	S16356	ovo protein - frui
23	34	85.0	1323	2	A24534	regulatory protein
24	33	82.5	138	2	T21792	hypothetical prote
25	33	82.5	201	2	I57505	zinc finger protei
26	33	82.5	223	2	F31201	GLI-related finger
27	33	82.5	225	2	D69090	conserved hypothet
28	33	82.5	233	2	D84797	hypothetical prote
29	33	82.5	240	2	S34619	probable fumarate
30	33	82.5	265	2	T33695	hypothetical prote
31	33	82.5	265	2	T15486	hypothetical prote
32	33	82.5	283	2	G96979	uncharacterized pr
33	33	82.5	293	2	T22385	hypothetical prote
34	33	82.5	319	2	T46469	hypothetical prote
35	33	82.5	395	2	D70441	oxido/reductase ir
36	33	82.5	395	2	T52423	C2H2 zinc finger p
37	33	82.5	427	2	A35659	krueppel-related p
38	33	82.5	501	1	F64462	polyferredoxin 2 -
39	33	82.5	544	2	T27444	hypothetical prote
40	33	82.5	555	2	I53869	zinc finger protei
41	33	82.5	577	2	B37057	integrin beta-6 ch
42	33	82.5	580	2	A46538	Ig heavy chain, se
43	33	82.5	626	2	F72256	Fe-hydrogenase, su
44	33	82.5	656	2	JC2005	integrin beta-5 ch
45	33	82.5	693	2	I37570	zinc finger protei

#### ALIGNMENTS

##### RESULT 1

G81260

UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diamino-pimelate ligase (EC 6.3.2.13)  
Cj1641 [imported] - Campylobacter jejuni (strain NCTC 11168)

C;Species: Campylobacter jejuni

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C;Accession: G81260

R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.;  
Chillingworth, T.; Davies, R.M.; Feltwell, T.; Holroyd, S.; Jagels, K.;

Karlyshev, A.; Moule, S.; Pallen, M.J.; Penn, C.W.; Quail, M.; Rajandream, M.A.;  
Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell, B.G.

Nature 403, 665-668, 2000

A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.

A;Reference number: A81250; MUID:20150912; PMID:10688204

A;Accession: G81260

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-427 <PAR>

A;Cross-references: GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB73629.1;  
PID:g6969058; GSPDB:GN00120; CJSP:Cj1641  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: murE; Cj1641  
C;Superfamily: UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase  
C;Keywords: ligase

Query Match 100.0%; Score 40; DB 2; Length 427;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 86 YKCGLC 91

#### RESULT 2

JH0500  
zinc finger protein 29 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 08-Dec-2000  
C;Accession: JH0500  
R;Denny, P.; Ashworth, A.  
Gene 106, 221-227, 1991  
A;Title: A zinc finger protein-encoding gene expressed in the post-meiotic phase  
of spermatogenesis.  
A;Reference number: JH0500; MUID:92039080; PMID:1937051  
A;Accession: JH0500  
A;Molecule type: mRNA  
A;Residues: 1-614 <DEN>  
A;Cross-references: GB:X55126; NID:g55470; PIDN:CAA38920.1; PID:g55471  
A;Experimental source: testis  
C;Keywords: DNA binding; zinc finger  
F;217-244/Region: zinc finger  
F;245-272/Region: zinc finger  
F;273-300/Region: zinc finger  
F;301-328/Region: zinc finger  
F;329-356/Region: zinc finger  
F;357-384/Region: zinc finger  
F;385-412/Region: zinc finger  
F;413-440/Region: zinc finger  
F;441-468/Region: zinc finger  
F;469-496/Region: zinc finger  
F;497-524/Region: zinc finger  
F;525-552/Region: zinc finger  
F;553-580/Region: zinc finger  
F;581-608/Region: zinc finger

Query Match 100.0%; Score 40; DB 2; Length 614;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 446 YKCGLC 451

# RESULT 3

S00647

finger protein - African clawed frog

C;Species: *Xenopus laevis* (African clawed frog)

C;Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 31-Dec-1993

C;Accession: S00647

R;Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.

EMBO J. 6, 3065-3070, 1987

A;Title: Xfin: an embryonic gene encoding a multifingered protein in *Xenopus*.

A;Reference number: S00647; MUID:88082679; PMID:2826129

A;Accession: S00647

A;Molecule type: mRNA

A;Residues: 1-1350 <RUI>

A;Cross-references: EMBL:X06021

A;Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator

C;Genetics:

A;Gene: fin

C;Keywords: DNA binding; zinc finger

Query Match 100.0%; Score 40; DB 2; Length 1350;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

|||||

Db 1044 YKCGLC 1049

# RESULT 4

B43284

zinc finger protein ZNF48 - human (fragment)

C;Species: *Homo sapiens* (man)

C;Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999

C;Accession: B43284

R;Lichter, P.; Bray, P.; Ried, T.; Dawid, I.B.; Ward, D.C.

Genomics 13, 999-1007, 1992

A;Title: Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile site regions of human chromosomes.

A;Reference number: A43284; MUID:92372070; PMID:1505991

A;Accession: B43284

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-57 <LIC>

A;Cross-references: GB:M88358; NID:g340451; PIDN:AAA61316.1; PID:g340452

A;Note: sequence extracted from NCBI backbone (NCBIN:111632, NCBIP:111633)

Query Match 92.5%; Score 37; DB 2; Length 57;

Best Local Similarity 83.3%; Pred. No. 16;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

||||:|

Db 15 YKCGVC 20

# RESULT 5

A95105

conserved hypothetical protein SP0909 [imported] - Streptococcus pneumoniae (strain TIGR4)

C;Species: Streptococcus pneumoniae

C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C;Accession: A95105

R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidelberg, J.; DeBoy, R.T.; Haft, D.H.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.; Kolonay, J.F.; Nelson, W.C.; Peterson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, E.; Khouri, H.; Wolf, A.M.; Utterback, T.R.; Hansen, C.L.; McDonald, L.A.; Feldblyum, T.V.; Angiuoli, S.; Dickinson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, D.A.; Hollingshead, S.K.; Fraser, C.M.

A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: A95105

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-149 <KUR>

A;Cross-references: GB:AE005672; PIDN:AAK75034.1; PID:g14972383; GSPDB:GN00164; TIGR:SP4SP0909

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP0909

Query Match	92.5%;	Score 37;	DB 2;	Length 149;
Best Local Similarity	83.3%;	Pred. No. 31;		
Matches	5;	Conservative	1;	Mismatches 0; Indels 0; Gaps 0;

Qy	1 YKGLC 6
	:
Db	132 YRCGLC 137

# RESULT 6

A97973

conserved hypothetical protein spr0809 [imported] - Streptococcus pneumoniae (strain R6)

C;Species: Streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C;Accession: A97973

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Estrem, S.; Fritz, L.; Fu, D.J.; Fuller, W.; Geringer, C.; Gilmour, R.; Glass, J.S.; Khoja, H.; Kraft, A.; LaGace, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; McHenney, M.; McLeaster, K.; Mundy, C.; Nicas, T.I.; Norris, F.H.; O'Gara, M.; Peery, R.; Robertson, G.T.; Rockey, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; Rostek Jr., P.R.; Skatrud, P.L.; Glass, J.I.

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: A97973

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-149 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAK99613.1; PID:g15458409; GSPDB:GN00174  
C;Genetics:  
A;Gene: spr0809

Query Match 92.5%; Score 37; DB 2; Length 149;  
Best Local Similarity 83.3%; Pred. No. 31;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|:||||  
Db 132 YRCGLC 137

RESULT 7

T25352

hypothetical protein T27C5.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T25352

R;Cummings, P.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z20020

A;Accession: T25352

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1638 <WIL>

A;Cross-references: EMBL:Z82058; PIDN:CAB04870.1; GSPDB:GN00023; CESP:T27C5.9

A;Experimental source: clone T27C5

C;Genetics:

A;Gene: CESP:T27C5.9

A;Map position: 5

A;Introns: 1432/2

Query Match 90.0%; Score 36; DB 2; Length 1638;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|:||||  
Db 159 YQCGLC 164

RESULT 8

C90165

conserved hypothetical protein [imported] - *Sulfolobus solfataricus*

C;Species: *Sulfolobus solfataricus*

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001

C;Accession: C90165

R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Weiher, C.C.Y.; Clausen, I.G.; Curtis, B.A.; De Moors, A.; Erauso, G.; Fletcher, C.; Gordon, P.M.K.; Heikamp-de Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.; Schenk, M.E.; Theriault, C.; Tolstrup, N.; Charlebois, R.L.; Doolittle, W.F.; Duguet, M.; Gaasterland, T.; Garrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A;Description: Sulfolobus solfataricus complete genome.

A;Reference number: A99139

A;Accession: C90165

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-582 <KUR>

A;Cross-references: GB:AE006641; NID:g13813374; PIDN:AAK40578.1; GSPDB:GN00155

C;Genetics:

A;Gene: SSO0235

C;Superfamily: conserved hypothetical protein MTH831

Query Match 87.5%; Score 35; DB 2; Length 582;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| ||||  
Db 122 YNCGLC 127

#### RESULT 9

S74069

hypothetical protein c0138 - Sulfolobus solfataricus

C;Species: Sulfolobus solfataricus

C;Date: 09-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 17-Mar-1999

C;Accession: S74069

R;Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.L.; Young, F.; Schenk, M.E.; Gaasterland, T.; Doolittle, W.F.; Ragan, M.A.; Charlebois, R.L.

Mol. Microbiol. 22, 175-191, 1996

A;Title: Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from Sulfolobus solfataricus P2.

A;Reference number: S73076; MUID:97055432; PMID:8899719

A;Accession: S74069

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1000 <SEN>

A;Cross-references: EMBL:Y08256; NID:g1707679; PID:e283937; PID:g1707768

A;Experimental source: strain P2

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996

Query Match 87.5%; Score 35; DB 2; Length 1000;  
Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
:|||||  
Db 573 HKCGLC 578

#### RESULT 10

T11614

probable poly(A)-specific ribonuclease (EC 3.1.13.4) - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 29-Oct-1999  
 C;Accession: T11614; S62448  
 R;Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.  
 submitted to the EMBL Data Library, October 1995  
 A;Reference number: Z17300  
 A;Accession: T11614  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1115 <BAR>  
 A;Cross-references: EMBL:Z54328; NID:g1009451; PID:g1009454  
 A;Experimental source: strain 972h(-)  
 R;Badcock, K.; Churcher, C.M.  
 submitted to the EMBL Data Library, October 1995  
 A;Reference number: S62445  
 A;Accession: S62448  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1115 <BAD>  
 A;Cross-references: EMBL:Z54328; NID:g1009451; PIDN:CAA91128.1; PID:g1009454  
 C;Genetics:  
 A;Map position: IL  
 A;Introns: 6/3; 39/3; 95/2; 164/3; 357/2; 793/1  
 A;Note: SPAC22G7.04  
 C;Keywords: hydrolase

Query Match 87.5%; Score 35; DB 2; Length 1115;  
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |:||:|  
 Db 640 YRCGIC 645

# RESULT 11

T22278

hypothetical protein F46B3.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000

C;Accession: T22278

R;Ainscough, R.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19541

A;Accession: T22278

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-186 <WIL>

A;Cross-references: EMBL:Z81540; PIDN:CAB04402.1; GSPDB:GN00023; CESP:F46B3.1

A;Experimental source: clone F46B3

C;Genetics:

A;Gene: CESP:F46B3.1

A;Map position: 5

A;Introns: 42/2; 79/2; 115/2; 155/2

C;Superfamily: Caenorhabditis elegans hypothetical protein F49F1.6

Query Match 85.0%; Score 34; DB 2; Length 186;  
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;



Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| ||||  
Db 46 YTCGLC 51

RESULT 12

T22279

hypothetical protein F46B3.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000

C;Accession: T22279

R;Ainscough, R.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19541

A;Accession: T22279

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-190 <WIL>

A;Cross-references: EMBL:Z81540; PIDN:CAB04403.1; GSPDB:GN00023; CESP:F46B3.2

A;Experimental source: clone F46B3

C;Genetics:

A;Gene: CESP:F46B3.2

A;Map position: 5

A;Introns: 94/2; 119/2; 159/2

C;Superfamily: Caenorhabditis elegans hypothetical protein F49F1.6

Query Match 85.0%; Score 34; DB 2; Length 190;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| ||||  
Db 61 YTCGLC 66

RESULT 13

T27824

hypothetical protein ZK287.6 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T27824

R;McMurray, A.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z20425

A;Accession: T27824

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-277 <WIL>

A;Cross-references: EMBL:Z70757; PIDN:CAA94802.1; GSPDB:GN00023; CESP:ZK287.6

A;Experimental source: clone ZK287

C;Genetics:

A;Gene: CESP:ZK287.6

A;Map position: 5

A;Introns: 37/2; 66/3; 144/1

Query Match 85.0%; Score 34; DB 2; Length 277;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
|:|:|  
Db 206 YECGIC 211

RESULT 14

F83044

nitrate-inducible formate dehydrogenase, beta subunit PA4811 [imported] -  
*Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 02-Aug-2002

C;Accession: F83044

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: F83044

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-309 <STO>

A;Cross-references: GB:AE004894; GB:AE004091; NID:g9951076; PIDN:AAG08197.1;

GSPDB:GN00131; PASP:PA4811

A;Experimental source: strain PA01

C;Superfamily: formate dehydrogenase, nitrate-inducible, beta chain; ferredoxin 2[4Fe-4S] homology

Query Match 85.0%; Score 34; DB 2; Length 309;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
||| ||  
Db 154 YKSLC 159

RESULT 15

C84797

hypothetical protein At2g37800 [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: C84797

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,

G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84797

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-396 <STO>

A;Cross-references: GB:AE002093; NID:g4895204; PIDN:AAD32791.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g37800

A;Map position: 2

Query Match 85.0%; Score 34; DB 2; Length 396;

Best Local Similarity 83.3%; Pred. No. 1.8e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

|||||

Db 71 YKCGQC 76

Search completed: March 5, 2004, 16:28:54

Job time : 1.37037 secs

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

```
Run on:      March  5, 2004, 16:22:54 ; Search time 3.24074 Seconds
              (without alignments)
              390.935 Million cell updates/sec
```

Title: US-10-057-890A-6  
Perfect score: 40  
Sequence: 1 YKCGLC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result		Query				Description
No.	Score	Match	Length	DB	ID	

1	40	100.0	6	14	US-10-057-890A-6	Sequence 6, Appli
2	40	100.0	138	14	US-10-057-890A-10	Sequence 10, Appl
3	40	100.0	157	14	US-10-057-890A-31	Sequence 31, Appl
4	38	95.0	357	15	US-10-104-047-2570	Sequence 2570, Ap
5	38	95.0	406	15	US-10-094-749-2523	Sequence 2523, Ap
6	38	95.0	519	15	US-10-094-749-1972	Sequence 1972, Ap
7	38	95.0	618	15	US-10-094-749-2479	Sequence 2479, Ap
8	38	95.0	670	15	US-10-108-260A-3103	Sequence 3103, Ap
9	38	95.0	714	15	US-10-108-260A-2908	Sequence 2908, Ap
10	37	92.5	193	15	US-10-374-780A-1990	Sequence 1990, Ap
11	37	92.5	522	15	US-10-108-260A-2767	Sequence 2767, Ap
12	37	92.5	799	15	US-10-104-047-2929	Sequence 2929, Ap
13	36	90.0	3067	10	US-09-949-029-18	Sequence 18, Appl
14	35	87.5	250	9	US-09-867-550-760	Sequence 760, App
15	35	87.5	313	9	US-09-800-729-196	Sequence 196, App
16	35	87.5	316	9	US-09-764-864-967	Sequence 967, App
17	34	85.0	23	9	US-09-785-632A-39	Sequence 39, Appl
18	34	85.0	23	14	US-10-223-765-39	Sequence 39, Appl
19	34	85.0	23	14	US-10-314-669-88	Sequence 88, Appl
20	34	85.0	308	9	US-09-764-864-927	Sequence 927, App
21	34	85.0	309	9	US-09-764-864-1366	Sequence 1366, Ap
22	34	85.0	327	15	US-10-108-260A-3050	Sequence 3050, Ap
23	34	85.0	366	10	US-09-949-029-96	Sequence 96, Appl
24	34	85.0	378	14	US-10-314-669-18	Sequence 18, Appl
25	34	85.0	530	10	US-09-372-348-11	Sequence 11, Appl
26	34	85.0	600	9	US-09-764-864-957	Sequence 957, App
27	34	85.0	683	15	US-10-104-047-2973	Sequence 2973, Ap
28	34	85.0	781	15	US-10-104-047-2937	Sequence 2937, Ap
29	34	85.0	1323	9	US-09-801-368-34	Sequence 34, Appl
30	33	82.5	23	14	US-10-314-669-77	Sequence 77, Appl
31	33	82.5	55	9	US-09-796-692-2505	Sequence 2505, Ap
32	33	82.5	55	14	US-10-040-862-2505	Sequence 2505, Ap
33	33	82.5	55	15	US-10-057-475B-2505	Sequence 2505, Ap
34	33	82.5	55	15	US-10-154-884B-2505	Sequence 2505, Ap
35	33	82.5	63	15	US-10-242-355-505	Sequence 505, App
36	33	82.5	79	14	US-10-029-386-28542	Sequence 28542, A
37	33	82.5	129	9	US-09-764-864-1352	Sequence 1352, Ap
38	33	82.5	213	14	US-10-314-669-12	Sequence 12, Appl
39	33	82.5	240	15	US-10-369-493-13535	Sequence 13535, A
40	33	82.5	240	15	US-10-369-493-18227	Sequence 18227, A
41	33	82.5	301	14	US-10-029-386-33634	Sequence 33634, A
42	33	82.5	334	9	US-09-764-864-908	Sequence 908, App
43	33	82.5	352	8	US-08-979-847-118	Sequence 118, App
44	33	82.5	352	14	US-10-114-104-118	Sequence 118, App
45	33	82.5	375	14	US-10-029-386-33792	Sequence 33792, A

#### ALIGNMENTS

RESULT 1  
 US-10-057-890A-6  
 ; Sequence 6, Application US/10057890A  
 ; Publication No. US20030044901A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Coleman, Timothy

; APPLICANT: Mansfield, Brian  
; TITLE OF INVENTION: Scaffold Fusion Polypeptides, Composition for Making the Same, and Methods  
; TITLE OF INVENTION: of Using the Same.  
; FILE REFERENCE: PF537  
; CURRENT APPLICATION NUMBER: US/10/057,890A  
; CURRENT FILING DATE: 2002-01-29  
; PRIOR APPLICATION NUMBER: 60/265,782  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: 60/265,858  
; PRIOR FILING DATE: 2001-01-31  
; NUMBER OF SEQ ID NOS: 32  
; SEQ ID NO 6  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-057-890A-6

Query Match 100.0%; Score 40; DB 14; Length 6;  
Best Local Similarity 100.0%; Pred. No. 7.2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 1 YKCGLC 6

RESULT 2

US-10-057-890A-10

; Sequence 10, Application US/10057890A  
; Publication No. US20030044901A1  
; GENERAL INFORMATION:  
; APPLICANT: Coleman, Timothy  
; APPLICANT: Mansfield, Brian  
; TITLE OF INVENTION: Scaffold Fusion Polypeptides, Composition for Making the Same, and Methods  
; TITLE OF INVENTION: of Using the Same.  
; FILE REFERENCE: PF537  
; CURRENT APPLICATION NUMBER: US/10/057,890A  
; CURRENT FILING DATE: 2002-01-29  
; PRIOR APPLICATION NUMBER: 60/265,782  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: 60/265,858  
; PRIOR FILING DATE: 2001-01-31  
; NUMBER OF SEQ ID NOS: 32  
; SEQ ID NO 10  
; LENGTH: 138  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-057-890A-10

Query Match 100.0%; Score 40; DB 14; Length 138;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||

Db

31 YKCGLC 36

RESULT 3

US-10-057-890A-31

; Sequence 31, Application US/10057890A

; Publication No. US20030044901A1

; GENERAL INFORMATION:

; APPLICANT: Coleman, Timothy

; APPLICANT: Mansfield, Brian

; TITLE OF INVENTION: Scaffold Fusion Polypeptides, Composition for Making the Same, and Methods

; TITLE OF INVENTION: of Using the Same.

; FILE REFERENCE: PF537

; CURRENT APPLICATION NUMBER: US/10/057,890A

; CURRENT FILING DATE: 2002-01-29

; PRIOR APPLICATION NUMBER: 60/265,782

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,858

; PRIOR FILING DATE: 2001-01-31

; NUMBER OF SEQ ID NOS: 32

; SEQ ID NO 31

; LENGTH: 157

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-057-890A-31

Query Match 100.0%; Score 40; DB 14; Length 157;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

|||||

Db 50 YKCGLC 55

RESULT 4

US-10-104-047-2570

; Sequence 2570, Application US/10104047

; Publication No. US20030236392A1

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: No. US20030236392A1 full length cDNA

; FILE REFERENCE: H1-A0105

; CURRENT APPLICATION NUMBER: US/10/104,047

; CURRENT FILING DATE: 2002-03-25

; PRIOR APPLICATION NUMBER:

; PRIOR FILING DATE:

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2570

; LENGTH: 357

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-104-047-2570

Query Match 95.0%; Score 38; DB 15; Length 357;

Best Local Similarity 83.3%; Pred. No. 86;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|  
Db 334 YKCGIC 339

RESULT 5

US-10-094-749-2523  
; Sequence 2523, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHIKO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2523  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2523

Query Match 95.0%; Score 38; DB 15; Length 406;  
Best Local Similarity 83.3%; Pred. No. 95;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|  
Db 383 YKCGIC 388

RESULT 6



US-10-094-749-1972  
 ; Sequence 1972, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SEKI, NAOHIKO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOYUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094,749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350,435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1972  
 ; LENGTH: 519  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-094-749-1972

Query Match 95.0%; Score 38; DB 15; Length 519;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 ||||:|  
 Db 435 YKCGMC 440

RESULT 7  
 US-10-094-749-2479  
 ; Sequence 2479, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI

```
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2479
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2479
```

```
Query Match          95.0%; Score 38; DB 15; Length 618;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches      5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 YKCGLC 6
        ||||:|
Db      220 YKCGIC 225
```

```
RESULT 8
US-10-108-260A-3103
; Sequence 3103, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3103
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3103
```

```
Query Match          95.0%; Score 38; DB 15; Length 670;
```

Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|  
Db 620 YKCGMC 625

RESULT 9

US-10-108-260A-2908  
; Sequence 2908, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2908  
; LENGTH: 714  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-2908

Query Match 95.0%; Score 38; DB 15; Length 714;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|  
Db 691 YKCGIC 696

RESULT 10

US-10-374-780A-1990  
; Sequence 1990, Application US/10374780A  
; Publication No. US20040019927A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, Bradley K  
; APPLICANT: Riechmann, Jose Luis  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Heard, Jacqueline E  
; APPLICANT: Haake, Volker  
; APPLICANT: Creelman, Robert A  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Adam, Luc J  
; APPLICANT: Reuber, T. Lynne  
; APPLICANT: Keddie, James  
; APPLICANT: Broun, Pierre E  
; APPLICANT: Pilgrim, Marsha L  
; APPLICANT: Dubell III, Arnold T  
; APPLICANT: Pineda, Omaira  
; APPLICANT: Yu, Guo-Liang  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS  
; FILE REFERENCE: MBI-0047 CIP

```

; CURRENT APPLICATION NUMBER: US/10/374,780A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/837,944
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 10/225,066
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,067
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,068
; PRIOR FILING DATE: 2002-08-09
; NUMBER OF SEQ ID NOS: 2906
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1990
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G351 Paralogous to G545
US-10-374-780A-1990

```

```

Query Match          92.5%; Score 37; DB 15; Length 193;
Best Local Similarity 83.3%; Pred. No. 74;
Matches      5; Conservative 1; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 YKCGLC 6
        ||||:|
Db      75 YKCGVC 80

```

```

RESULT 11
US-10-108-260A-2767
; Sequence 2767, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2767
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-2767

```

Query Match 92.5%; Score 37; DB 15; Length 522;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|:||||  
Db 412 YRCGLC 417

RESULT 12

US-10-104-047-2929  
; Sequence 2929, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2929  
; LENGTH: 799  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2929

Query Match 92.5%; Score 37; DB 15; Length 799;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|  
Db 632 YKCGVC 637

RESULT 13

US-09-949-029-18  
; Sequence 18, Application US/09949029  
; Publication No. US20030134278A1  
; GENERAL INFORMATION:  
; APPLICANT: Karpen, G.H.  
; APPLICANT: Dobie, K.W.  
; APPLICANT: Kennedy, C.D.  
; APPLICANT: Velasco, V.M.  
; APPLICANT: McGrath, T.L.  
; APPLICANT: Weko, J.  
; APPLICANT: Patterson, R.W.  
; TITLE OF INVENTION: Identification of chromosome inheritance modifiers in  
Drosophila  
; TITLE OF INVENTION: melanogaster  
; FILE REFERENCE: 1211.015US1  
; CURRENT APPLICATION NUMBER: US/09/949,029  
; CURRENT FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: US 60/231,178

; PRIOR FILING DATE: 2000-09-07  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 3067  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-949-029-18

Query Match 90.0%; Score 36; DB 10; Length 3067;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|:||||  
Db 2813 YQCGLC 2818

RESULT 14

US-09-867-550-760  
; Sequence 760, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1e1 Polynucleotides from Atherogenic  
Cells and Polypeptides Encoded  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 760  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-867-550-760

Query Match 87.5%; Score 35; DB 9; Length 250;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 82 YKCGAC 87

RESULT 15

US-09-800-729-196  
; Sequence 196, Application US/09800729  
; Patent No. US20020068319A1

; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: PZ044P1  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 196  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-729-196

Query Match 87.5%; Score 35; DB 9; Length 313;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
| | | | |  
Db 208 YKCGAC 213

Search completed: March 5, 2004, 16:33:42  
Job time : 3.24074 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:15:44 ; Search time 3.98148 Seconds  
(without alignments)  
475.479 Million cell updates/sec

Title: US-10-057-890A-6  
Perfect score: 40  
Sequence: 1 YKCGLC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result		Query				
No.	Score	Match	Length	DB	ID	Description



1	40	100.0	127	4	Q9BTS6	Q9bts6 homo sapien
2	40	100.0	262	5	Q86S59	Q86s59 caenorhabdi
3	40	100.0	366	11	Q8CIM2	Q8cim2 mus musculu
4	40	100.0	433	5	Q95X98	Q95x98 caenorhabdi
5	40	100.0	601	11	Q8BJ07	Q8bj07 mus musculu
6	38	95.0	190	5	Q8SW43	Q8sw43 encephalito
7	38	95.0	357	4	Q96JW6	Q96jw6 homo sapien
8	38	95.0	406	4	Q96MV3	Q96mv3 homo sapien
9	38	95.0	563	11	Q921H7	Q921h7 mus musculu
10	38	95.0	591	11	Q8R0V0	Q8r0v0 mus musculu
11	38	95.0	618	4	Q96MX3	Q96mx3 homo sapien
12	38	95.0	652	4	Q9NS42	Q9ns42 homo sapien
13	38	95.0	714	4	Q8TE95	Q8te95 homo sapien
14	38	95.0	769	4	Q8IZD3	Q8izd3 homo sapien
15	38	95.0	837	5	Q9NAS7	Q9nas7 anopheles g
16	37	92.5	57	4	Q15920	Q15920 homo sapien
17	37	92.5	149	16	Q97RB5	Q97rb5 streptococc
18	37	92.5	149	16	Q8DQ78	Q8dq78 streptococc
19	37	92.5	193	10	Q8L8P4	Q8l8p4 arabidopsis
20	37	92.5	193	10	Q9SSW0	Q9ssw0 arabidopsis
21	37	92.5	228	10	Q8H5Y7	Q8h5y7 oryza sativ
22	37	92.5	337	11	Q8BG98	Q8bg98 mus musculu
23	37	92.5	451	11	Q8BIK0	Q8bik0 mus musculu
24	37	92.5	522	4	Q8N9F8	Q8n9f8 homo sapien
25	37	92.5	536	11	Q80Y34	Q80y34 mus musculu
26	37	92.5	538	5	Q9VGG0	Q9vvg0 drosophila
27	37	92.5	613	4	Q7Z7L9	Q7z7l9 homo sapien
28	37	92.5	626	4	Q9BX82	Q9bx82 homo sapien
29	37	92.5	750	13	Q7ZYB2	Q7zyb2 xenopus lae
30	37	92.5	751	4	Q9NS43	Q9ns43 homo sapien
31	37	92.5	799	4	Q86WZ6	Q86wz6 homo sapien
32	37	92.5	1004	5	Q8SZM4	Q8szm4 drosophila
33	36	90.0	155	10	Q8H7J9	Q8h7j9 oryza sativ
34	36	90.0	348	10	Q8GXX6	Q8gxx6 arabidopsis
35	36	90.0	348	13	O93384	O93384 oncorhynchu
36	36	90.0	360	11	Q8VHM2	Q8vhm2 mus musculu
37	36	90.0	360	11	O70600	O70600 rattus norv
38	36	90.0	361	4	Q8WVI4	Q8wvi4 homo sapien
39	36	90.0	361	4	Q8WXG1	Q8wxg1 homo sapien
40	36	90.0	362	6	Q9MZU4	Q9mzu4 sus scrofa
41	36	90.0	362	11	Q9JHD4	Q9jhd4 mus musculu
42	36	90.0	362	11	Q8CBB9	Q8cbb9 mus musculu
43	36	90.0	388	5	Q8SUH6	Q8suh6 encephalito
44	36	90.0	523	11	Q80TE5	Q80te5 mus musculu
45	36	90.0	581	11	Q8C7A0	Q8c7a0 mus musculu

# ALIGNMENTS

## RESULT 1

### Q9BTS6

ID	Q9BTS6	PRELIMINARY;	PRT;	127 AA.
AC	Q9BTS6;			
DT	01-JUN-2001	(TrEMBLrel. 17, Created)		
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)		

DE Hypothetical protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC003386; AAH03386.1; -.  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 127 AA; 14474 MW; 24C29EA7A20FABD1 CRC64;

Query Match 100.0%; Score 40; DB 4; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||  
 Db 65 YKCGLC 70

## RESULT 2

Q86S59

ID Q86S59 PRELIMINARY; PRT; 262 AA.  
 AC Q86S59;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein Y37F4.6b.  
 GN Y37F4.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AC098856; AAO21408.1; -.  
 DR WormPep; Y37F4.6b; CE33214.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW Hypothetical protein.

SQ SEQUENCE 262 AA; 31141 MW; 8ACF36F0E6420A2E CRC64;

Query Match 100.0%; Score 40; DB 5; Length 262;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 93 YKCGLC 98

RESULT 3

Q8CIM2

ID Q8CIM2 PRELIMINARY; PRT; 366 AA.  
AC Q8CIM2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to zinc finger protein 29.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CZECH II;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC023282; AAH23282.1; -.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF00096; zf-C2H2; 9.  
DR PRINTS; PR00048; ZINCFINGER.  
DR SMART; SM00355; ZnF\_C2H2; 9.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 10.  
SQ SEQUENCE 366 AA; 41293 MW; 3C4BE0C8A73700B6 CRC64;

Query Match 100.0%; Score 40; DB 11; Length 366;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
|||||  
Db 117 YKCGLC 122

RESULT 4

Q95X98

ID Q95X98 PRELIMINARY; PRT; 433 AA.  
AC Q95X98;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN Y37F4.6.  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AC098856; AAL27253.1; -.  
 DR WormPep; Y37F4.6a; CE29854.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 1.  
 DR SMART; SM00355; Znf\_C2H2; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 433 AA; 49833 MW; 9F373420CBDE834D CRC64;

Query Match 100.0%; Score 40; DB 5; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||  
 Db 93 YKCGLC 98

# RESULT 5

Q8BJ07

ID Q8BJ07 PRELIMINARY; PRT; 601 AA.  
 AC Q8BJ07;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Zinc finger protein 3l.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).

DR EMBL; AK047820; BAC33164.1; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR InterPro; IPR007086; Znf\_C2H2\_sub.  
 DR Pfam; PF00096; zf-C2H2; 9.  
 DR PRINTS; PR00048; ZINCFINGER.  
 DR SMART; SM00717; SANT; 1.  
 DR SMART; SM00355; ZnF\_C2H2; 9.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 10.  
 SQ SEQUENCE 601 AA; 67542 MW; 90B65BF55B117C08 CRC64;

Query Match 100.0%; Score 40; DB 11; Length 601;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||  
 Db 352 YKCGLC 357

# RESULT 6

Q8SW43

ID Q8SW43 PRELIMINARY; PRT; 190 AA.  
 AC Q8SW43;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein ECU03\_0790.  
 GN ECU03\_0790.  
 OS Encephalitozoon cuniculi.  
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.  
 OX NCBI\_TaxID=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806;  
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi.";  
 RL Nature 414:450-453(2001).  
 DR EMBL; AL590443; CAD26223.1; -.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 6.  
 DR SMART; SM00355; ZnF\_C2H2; 5.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.

KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 190 AA; 22245 MW; C2006BC95F3AA5B1 CRC64;

Query Match 95.0%; Score 38; DB 5; Length 190;  
Best Local Similarity 83.3%; Pred. No. 14;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|  
Db 85 YKCGIC 90

#### RESULT 7

Q96JW6

ID Q96JW6 PRELIMINARY; PRT; 357 AA.  
AC Q96JW6;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein FLJ14928.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; AK027834; BAB55400.1; -.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF00096; zf-C2H2; 12.  
DR PRINTS; PR00048; ZINCFINGER.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00355; ZnF\_C2H2; 13.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 13.  
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;  
KW Zinc-finger.  
SQ SEQUENCE 357 AA; 40426 MW; 541DBE6D009BBC04 CRC64;

Query Match 95.0%; Score 38; DB 4; Length 357;  
Best Local Similarity 83.3%; Pred. No. 23;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
||||:|

## RESULT 8

Q96MV3

ID Q96MV3 PRELIMINARY; PRT; 406 AA.  
 AC Q96MV3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ31827.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK056389; BAB71174.1; -.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 10.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; ZnF\_C2H2; 10.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 10.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 12.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 406 AA; 46686 MW; 1BAB9BC5DF908886 CRC64;

Query Match 95.0%; Score 38; DB 4; Length 406;

Best Local Similarity 83.3%; Pred. No. 26;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6

||||:

Db 383 YKCGIC 388

## RESULT 9

Q921H7

ID Q921H7 PRELIMINARY; PRT; 563 AA.  
 AC Q921H7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN BC026401.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC012403; AAH12403.1; -.  
 DR MGD; MGI:2384725; BC026401.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 12.  
 DR SMART; SM00355; Znf\_C2H2; 12.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 12.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 563 AA; 61463 MW; D446AD9E8ED7C45F CRC64;

Query Match 95.0%; Score 38; DB 11; Length 563;  
 Best Local Similarity 83.3%; Pred. No. 34;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKCGLC 6  
 ||||:|  
 Db 163 YKCGIC 168

# RESULT 10

## Q8ROV0

ID Q8ROV0 PRELIMINARY; PRT; 591 AA.  
 AC Q8ROV0;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein (Hypothetical zinc finger).  
 GN BC026401.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";



RL Nature 420:563-573(2002).  
 DR EMBL; BC026401; AAH26401.1; -.  
 DR EMBL; AK049208; BAC33610.1; -.  
 DR MGD; MGI:2384725; BC026401.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 12.  
 DR SMART; SM00355; ZnF\_C2H2; 12.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 12.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 591 AA; 64599 MW; EE070D2692FA4A73 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 591;  
 Best Local Similarity 83.3%; Pred. No. 35;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 ||||:|  
 Db 191 YKCGIC 196

# RESULT 11

Q96MX3

ID Q96MX3 PRELIMINARY; PRT; 618 AA.  
 AC Q96MX3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein FLJ31751.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK056313; BAB71146.1; -.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 12.  
 DR SMART; SM00355; ZnF\_C2H2; 12.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 12.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 618 AA; 67832 MW; 28C0F13E54773353 CRC64;

Query Match 95.0%; Score 38; DB 4; Length 618;  
 Best Local Similarity 83.3%; Pred. No. 36;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
| | | | : |  
Db 220 YKCGIC 225

RESULT 12

Q9NS42

ID Q9NS42 PRELIMINARY; PRT; 652 AA.  
AC Q9NS42;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical zinc finger-like protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kodoyianni V., Ge Y., Severin J., Krummel G.K., Grable L.,  
RA Kvikstad E., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;  
RT "Sequence analysis of a 1Mb region in 19ql3.2 containing a zinc finger  
RT gene cluster."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; AC074331; AAF88106.1; -.  
DR HSSP; P08045; 1ZNF.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0008237; F:metallopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF00096; zf-C2H2; 8.  
DR PRINTS; PR00048; ZINCFINGER.  
DR ProDom; PD000003; Znf\_C2H2; 7.  
DR SMART; SM00355; ZnF\_C2H2; 7.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 8.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.  
SQ SEQUENCE 652 AA; 72644 MW; 3C9F250594170178 CRC64;

Query Match 95.0%; Score 38; DB 4; Length 652;  
Best Local Similarity 83.3%; Pred. No. 38;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
| | | | : |  
Db 602 YKCGMC 607

RESULT 13

Q8TE95

ID Q8TE95 PRELIMINARY; PRT; 714 AA.

AC Q8TE95;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ23765.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK074345; BAB85057.1; -.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 21.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; ZnF\_C2H2; 21.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 21.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 23.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 714 AA; 81668 MW; EFD0E080DE78FC06 CRC64;

Query Match 95.0%; Score 38; DB 4; Length 714;  
 Best Local Similarity 83.3%; Pred. No. 41;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 ||||:|  
 Db 691 YKCGIC 696

#### RESULT 14

Q8IZD3

ID Q8IZD3 PRELIMINARY; PRT; 769 AA.  
 AC Q8IZD3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Zinc finger protein 41-like protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Guo J.H., Yu L.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AY147015; AAN61169.1; -.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF01352; KRAB; 1.  
 DR Pfam; PF00096; zf-C2H2; 21.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; ZnF\_C2H2; 21.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 21.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 23.  
 SQ SEQUENCE 769 AA; 88000 MW; CC670B86FFCABBFD CRC64;

Query Match 95.0%; Score 38; DB 4; Length 769;  
 Best Local Similarity 83.3%; Pred. No. 43;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 ||||:|  
 Db 746 YKCGIC 751

# RESULT 15

## Q9NAS7

ID Q9NAS7 PRELIMINARY; PRT; 837 AA.  
 AC Q9NAS7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Integrin beta subunit precursor.  
 GN BINT.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 OX NCBI\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=G3;  
 RX MEDLINE=21331701; PubMed=11437913;  
 RA Machairaki V., Lycett G., Blass C., Louis C.;  
 RT "Beta-integrin of Anopheles gambiae: mRNA cloning and analysis of  
 RT structure and expression.";  
 RL Insect Mol. Biol. 10:217-223(2001).  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE  
 CC BONDS (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.  
 DR EMBL; AJ292755; CAC00630.1; -.  
 DR HSSP; P05106; 1JV2.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0008305; C:integrin complex; IEA.  
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR GO; GO:0007275; P:development; IEA.  
 DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.

DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR002369; Integrin\_B.  
 DR InterPro; IPR001169; Integrin\_beta\_C.  
 DR InterPro; IPR003659; Plexin-like.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00362; integrin\_B; 1.  
 DR PRINTS; PR01186; INTEGRINB.  
 DR ProDom; PD001811; Integrin\_B; 1.  
 DR SMART; SM00187; INB; 1.  
 DR SMART; SM00423; PSI; 1.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00243; INTEGRIN\_BETA; 3.  
 KW Cell adhesion; Glycoprotein; Integrin; Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 837 BETA INTEGRIN SUBUNIT.  
 SQ SEQUENCE 837 AA; 92791 MW; 30F1F339D6924D78 CRC64;

Query Match 95.0%; Score 38; DB 5; Length 837;  
 Best Local Similarity 83.3%; Pred. No. 46;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 ||||:|  
 Db 518 YKCGIC 523

Search completed: March 5, 2004, 16:27:28  
 Job time : 4.98148 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:15:14 ; Search time 0.814815 Seconds  
(without alignments)  
383.426 Million cell updates/sec

Title: US-10-057-890A-6  
Perfect score: 40  
Sequence: 1 YKCGLC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	100.0	427	1	MURE_CAMJE	O69290 campylobact
2	40	100.0	614	1	ZF29_MOUSE	Q07230 mus musculu
3	40	100.0	1350	1	XFIN_XENLA	P08045 xenopus lae
4	37	92.5	522	1	Z454_HUMAN	Q8n9f8 homo sapien
5	37	92.5	626	1	Z471_HUMAN	Q9bx82 homo sapien
6	36	90.0	581	1	Z319_MOUSE	Q9err8 mus musculu
7	36	90.0	582	1	Z319_HUMAN	Q9p2f9 homo sapien
8	35	87.5	682	1	ZN45_HUMAN	Q02386 homo sapien
9	35	87.5	722	1	Z219_HUMAN	Q9p2y4 homo sapien
10	35	87.5	1115	1	YAA4_SCHPO	Q09798 schizosacch
11	34	85.0	235	1	OVO1_HUMAN	O14753 homo sapien
12	34	85.0	267	1	OVO1_MOUSE	Q9wtj2 mus musculu
13	34	85.0	352	1	Z185_MOUSE	Q62394 mus musculu
14	34	85.0	457	1	Z185_HUMAN	O15231 homo sapien
15	34	85.0	472	1	LI26_CAEEL	Q27355 caenorhabdi
16	34	85.0	495	1	MLP2_DROME	Q24400 drosophila
17	34	85.0	506	1	Y619_METJA	Q58036 methanococc

18	34	85.0	514	1	ZBT2_HUMAN	Q8n680	homo sapien
19	34	85.0	605	1	TC17_HUMAN	O60765	homo sapien
20	34	85.0	798	1	ITB5_MOUSE	O70309	mus musculu
21	34	85.0	1028	1	OVO_DROME	P51521	drosophila
22	34	85.0	1031	1	Z445_HUMAN	P59923	homo sapien
23	34	85.0	1323	1	ADR1_YEAST	P07248	saccharomyc
24	33	82.5	201	1	Z239_MOUSE	P24399	mus musculu
25	33	82.5	273	1	ZN80_GORGO	P51503	gorilla gor
26	33	82.5	293	1	ZN80_MACMU	P51505	macaca mula
27	33	82.5	376	1	GLI4_HUMAN	P10075	homo sapien
28	33	82.5	407	1	OZF_MOUSE	Q62513	mus musculu
29	33	82.5	407	1	OZF_RAT	Q62981	rattus norv
30	33	82.5	408	1	Y51B_METJA	P81293	methanococc
31	33	82.5	458	1	Z239_HUMAN	Q16600	homo sapien
32	33	82.5	501	1	YD03_METJA	Q58699	methanococc
33	33	82.5	577	1	ITB6_CAVPO	P18563	cavia porce
34	33	82.5	634	1	Z297_HUMAN	O15209	homo sapien
35	33	82.5	645	1	Z235_MOUSE	Q61116	mus musculu
36	33	82.5	655	1	ITB5_PAPCY	Q07441	papio cynoc
37	33	82.5	697	1	Z450_HUMAN	O43167	homo sapien
38	33	82.5	698	1	Z234_HUMAN	Q14588	homo sapien
39	33	82.5	744	1	YJ62_HUMAN	Q8tf39	homo sapien
40	33	82.5	744	1	Z366_HUMAN	Q8n895	homo sapien
41	33	82.5	799	1	ITB5_HUMAN	P18084	homo sapien
42	33	82.5	803	1	Z226_HUMAN	Q9nyt6	homo sapien
43	33	82.5	813	1	CAN7_HUMAN	Q9y6w3	homo sapien
44	33	82.5	813	1	CAN7_MOUSE	Q9rls8	mus musculu
45	33	82.5	881	1	PHFE_MOUSE	Q9d4h9	mus musculu

# ALIGNMENTS

## RESULT 1

### MURE\_CAMJE

ID MURE\_CAMJE STANDARD; PRT; 427 AA.  
AC O69290; Q9PM35;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase  
DE (EC 6.3.2.13) (UDP-N-acetylmuramyl-tripeptide synthetase) (Meso-  
DE diaminopimelate-adding enzyme) (UDP-MurNAc-tripeptide synthetase).  
GN MURE OR CJ1641.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;

RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*  
 RT reveals hypervariable sequences."  
 RL Nature 403:665-668(2000).  
 RN [2]  
 RP SEQUENCE OF 1-326 FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RA Griffiths P.L., Connerton I.F.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Cell wall formation. Diaminopimelic acid adding enzyme  
 CC (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-  
 CC glutamate + meso-2,6-diaminoheptanedioate = ADP + phosphate + UDP-  
 CC N-acetylmuramoyl-L-alanyl-D-gamma-glutamyl-meso-2,6-diamino-  
 CC heptanedioate.  
 CC -!- PATHWAY: Peptidoglycan biosynthesis.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: Belongs to the murCDEF family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL139079; CAB73629.1; -.  
 DR EMBL; Y16882; CAA76495.1; ALT\_INIT.  
 DR PIR; G81260; G81260.  
 DR HAMAP; MF\_00208; -; 1.  
 DR InterPro; IPR000713; Mur\_ligase.  
 DR InterPro; IPR004101; Mur\_ligase\_C.  
 DR InterPro; IPR005761; MurE.  
 DR Pfam; PF01225; Mur\_ligase; 1.  
 DR Pfam; PF02875; Mur\_ligase\_C; 1.  
 DR TIGRFAMs; TIGR01085; murE; 1.  
 KW Peptidoglycan synthesis; Cell wall; Cell division; Ligase;  
 KW ATP-binding; Complete proteome.  
 FT NP\_BIND 66 72 ATP (POTENTIAL).  
 FT CONFLICT 1 3 MKL -> EIK (IN REF. 2).  
 FT CONFLICT 39 39 K -> G (IN REF. 2).  
 FT CONFLICT 54 54 K -> E (IN REF. 2).  
 FT CONFLICT 121 121 Q -> L (IN REF. 2).  
 FT CONFLICT 125 125 Q -> P (IN REF. 2).  
 FT CONFLICT 134 134 E -> G (IN REF. 2).  
 FT CONFLICT 142 142 Q -> P (IN REF. 2).  
 FT CONFLICT 145 145 I -> S (IN REF. 2).  
 FT CONFLICT 166 166 H -> R (IN REF. 2).  
 FT CONFLICT 170 170 E -> G (IN REF. 2).  
 SQ SEQUENCE 427 AA; 47721 MW; FADBF1E67471BB24 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 427;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||



## RESULT 2

## ZF29\_MOUSE

ID ZF29\_MOUSE STANDARD; PRT; 614 AA.  
AC Q07230;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Zinc finger protein 29 (Zfp-29).  
GN ZFP29 OR ZFP-29.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=92039080; PubMed=1937051;  
RA Denny P., Ashworth A.;  
RT "A zinc finger protein-encoding gene expressed in the post-meiotic  
RT phase of spermatogenesis.";  
RL Gene 106:221-227(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Olfactory epithelium;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION DURING  
CC THE POST-MEIOTIC STAGES OF SPERMATOGENESIS.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: IN THE ADULT, PREDOMINANTLY FOUND IN  
CC SPERMATIDS. ALSO PRESENT IN THE EMBYRO.  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: Contains 1 SCAN box domain.  
CC -!- SIMILARITY: Contains 14 C2H2-type zinc fingers.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55126; CAA38920.1; -.
DR EMBL; BC046961; AAH46961.1; -.
DR PIR; JH0500; JH0500.
DR HSSP; P08047; 1SP2.
DR MGD; MGI:99176; Zfp29.
DR InterPro; IPR003309; Treg_SCAN.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF02023; SCAN; 1.
DR Pfam; PF00096; zf-C2H2; 14.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 13.
DR SMART; SM00431; LER; 1.
DR SMART; SM00355; ZnF_C2H2; 14.
DR PROSITE; PS50804; SCAN_BOX; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat;
KW Developmental protein; Transcription regulation; Spermatogenesis.
FT DOMAIN 69 127 SCAN_BOX.
FT ZN_FING 222 244 C2H2-TYPE 1.
FT ZN_FING 250 272 C2H2-TYPE 2.
FT ZN_FING 278 300 C2H2-TYPE 3.
FT ZN_FING 306 328 C2H2-TYPE 4.
FT ZN_FING 334 356 C2H2-TYPE 5.
FT ZN_FING 362 384 C2H2-TYPE 6.
FT ZN_FING 390 412 C2H2-TYPE 7.
FT ZN_FING 418 440 C2H2-TYPE 8.
FT ZN_FING 446 468 C2H2-TYPE 9.
FT ZN_FING 474 496 C2H2-TYPE 10.
FT ZN_FING 502 524 C2H2-TYPE 11.
FT ZN_FING 530 552 C2H2-TYPE 12.
FT ZN_FING 558 580 C2H2-TYPE 13.
FT ZN_FING 586 608 C2H2-TYPE 14.
SQ SEQUENCE 614 AA; 68714 MW; 738156A0A9413DF7 CRC64;

```

```

Query Match 100.0%; Score 40; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 YKCGLC 6
    |||||
Db 446 YKCGLC 451

```

```

RESULT 3
XFIN_XENLA
ID XFIN_XENLA STANDARD; PRT; 1350 AA.

```

AC P08045;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein Xfin.  
 GN XFIN.  
 OS *Xenopus laevis* (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; *Xenopus*.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=88082679; PubMed=2826129;  
 RA Ruiz i Altaba A., Perry-O'Keefe H., Melton D.A.;  
 RT "Xfin: an embryonic gene encoding a multifingered protein in  
 RT *Xenopus*.";  
 RL EMBO J. 6:3065-3070(1987).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94021366; PubMed=7692399;  
 RA Andreazzoli M., de Lucchini S., Costa M., Barsacchi G.;  
 RT "RNA binding properties and evolutionary conservation of the *Xenopus*  
 RT multifinger protein Xfin.";  
 RL Nucleic Acids Res. 21:4218-4225(1993).  
 RN [3]  
 RP STRUCTURE BY NMR OF FINGER 31.  
 RX MEDLINE=89346749; PubMed=2503871;  
 RA Lee M.S., Gippert G.P., Soman K.V., Case D.A., Wright P.E.;  
 RT "Three-dimensional solution structure of a single zinc finger DNA-  
 RT binding domain.";  
 RL Science 245:635-637(1989).  
 RN [4]  
 RP STRUCTURE BY NMR OF A FINGER.  
 RX MEDLINE=89378224; PubMed=2506074;  
 RA Lee M.S., Cavanagh J., Wright P.E.;  
 RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc  
 RT finger from Xfin. Sequential resonance assignments and secondary  
 RT structure.";  
 RL FEBS Lett. 254:159-164(1989).  
 CC -!- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL  
 CC REGULATION PROCESSES.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH  
 CC AS NEURAL RETINA CONES.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND  
 CC EMBRYOGENESIS.  
 CC -!- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.  
 CC -!- PTM: Phosphorylated.  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -!- SIMILARITY: Contains 1 KRAB domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X06021; CAA29425.1; -.

DR PDB; 1ZNF; 15-OCT-91.

DR InterPro; IPR001909; KRAB.

DR InterPro; IPR007087; Znf\_C2H2.

DR Pfam; PF01352; KRAB; 1.

DR Pfam; PF00096; zf-C2H2; 36.

DR ProDom; PD000003; Znf\_C2H2; 20.

DR SMART; SM00349; KRAB; 1.

DR SMART; SM00355; Znf\_C2H2; 35.

DR PROSITE; PS50805; KRAB; 1.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 35.

DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 37.

KW Zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;

KW Phosphorylation.

FT	DOMAIN	1	58	KRAB.
FT	ZN_FING	108	130	C2H2-TYPE.
FT	ZN_FING	136	158	C2H2-TYPE.
FT	ZN_FING	164	186	C2H2-TYPE.
FT	ZN_FING	192	214	C2H2-TYPE.
FT	ZN_FING	220	242	C2H2-TYPE.
FT	ZN_FING	248	270	C2H2-TYPE.
FT	ZN_FING	276	298	C2H2-TYPE.
FT	ZN_FING	326	348	C2H2-TYPE.
FT	ZN_FING	354	376	C2H2-TYPE.
FT	ZN_FING	382	404	C2H2-TYPE.
FT	ZN_FING	410	432	C2H2-TYPE.
FT	ZN_FING	438	460	C2H2-TYPE.
FT	ZN_FING	466	488	C2H2-TYPE.
FT	ZN_FING	503	525	C2H2-TYPE.
FT	ZN_FING	531	553	C2H2-TYPE.
FT	ZN_FING	559	581	C2H2-TYPE.
FT	ZN_FING	587	609	C2H2-TYPE.
FT	ZN_FING	615	637	C2H2-TYPE.
FT	ZN_FING	643	665	C2H2-TYPE.
FT	ZN_FING	671	693	C2H2-TYPE.
FT	ZN_FING	699	721	C2H2-TYPE.
FT	ZN_FING	750	772	C2H2-TYPE.
FT	ZN_FING	778	800	C2H2-TYPE.
FT	ZN_FING	806	828	C2H2-TYPE.
FT	ZN_FING	834	856	C2H2-TYPE.
FT	ZN_FING	862	884	C2H2-TYPE.
FT	ZN_FING	890	912	C2H2-TYPE.
FT	ZN_FING	918	940	C2H2-TYPE.
FT	ZN_FING	988	1010	C2H2-TYPE.
FT	ZN_FING	1016	1038	C2H2-TYPE.
FT	ZN_FING	1044	1066	C2H2-TYPE.
FT	ZN_FING	1136	1158	C2H2-TYPE.
FT	ZN_FING	1164	1186	C2H2-TYPE.
FT	ZN_FING	1192	1214	C2H2-TYPE.
FT	ZN_FING	1220	1242	C2H2-TYPE.
FT	ZN_FING	1248	1270	C2H2-TYPE.
FT	ZN_FING	1276	1298	C2H2-TYPE.

FT STRAND 1045 1045  
 FT STRAND 1052 1052  
 FT HELIX 1056 1062  
 FT HELIX 1063 1065  
 FT TURN 1066 1066  
 SQ SEQUENCE 1350 AA; 155804 MW; 27F10AB0851E0AD8 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 1350;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 |||||  
 Db 1044 YKCGLC 1049

#### RESULT 4

Z454\_HUMAN

ID Z454\_HUMAN STANDARD; PRT; 522 AA.

AC Q8N9F8;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Zinc finger protein 454.

GN ZNF454.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,

RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,

RA Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,

RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,

RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,

RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T.,

RA Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K.,

RA Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y.,

RA Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K.,

RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: May be involved in transcriptional regulation.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC -!- SIMILARITY: Contains 1 KRAB domain.

CC -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AK094763; BAC04418.1; -.  
 DR Genew; HGNC:21200; ZNF454.  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR InterPro; IPR007086; Znf\_C2H2\_sub.  
 DR Pfam; PF01352; KRAB; 1.  
 DR Pfam; PF00096; zf-C2H2; 12.  
 DR PRINTS; PR00048; ZINCFINGER.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; Znf\_C2H2; 12.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 12.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat.  
 FT DOMAIN 14 84 KRAB.  
 FT ZN\_FING 178 200 C2H2-TYPE 1.  
 FT ZN\_FING 216 238 C2H2-TYPE 2.  
 FT ZN\_FING 244 266 C2H2-TYPE 3.  
 FT ZN\_FING 272 294 C2H2-TYPE 4.  
 FT ZN\_FING 300 322 C2H2-TYPE 5.  
 FT ZN\_FING 328 350 C2H2-TYPE 6.  
 FT ZN\_FING 356 378 C2H2-TYPE 7.  
 FT ZN\_FING 384 406 C2H2-TYPE 8.  
 FT ZN\_FING 412 434 C2H2-TYPE 9.  
 FT ZN\_FING 440 462 C2H2-TYPE 10.  
 FT ZN\_FING 468 490 C2H2-TYPE 11.  
 FT ZN\_FING 496 518 C2H2-TYPE 12.  
 SQ SEQUENCE 522 AA; 60068 MW; 9E633DC19056C405 CRC64;

Query Match 92.5%; Score 37; DB 1; Length 522;  
 Best Local Similarity 83.3%; Pred. No. 15;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 |:||||  
 Db 412 YRCGLC 417

# RESULT 5

## Z471\_HUMAN

ID Z471\_HUMAN STANDARD; PRT; 626 AA.  
 AC Q9BX82; O75260; Q8N3V1; Q9P2F1;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Zinc finger protein 471 (EZFIT-related protein 1).  
 GN ZNF471 OR ERP1 OR KIAA1396.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;  
 RA Matakai C., Murakami T., Umetani M., Wada Y., Hamakubo T., Kodama T.;  
 RT "EZFIT-related protein 1.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Koehrer K., Beyer A., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 76-626 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI.  
 RT The complete sequences of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:65-73(2000).  
 RN [4]  
 RP SEQUENCE OF 87-626 FROM N.A.  
 RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,  
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,  
 RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,  
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
 RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,  
 RA Kobayashi A., Olsen A.S., Carrano A.V.;  
 RT "Sequence analysis of a 2 Mb region containing a zinc finger (ZNF)  
 RT gene cluster in 19q13.4.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: May be involved in transcriptional regulation.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -!- SIMILARITY: Contains 1 KRAB domain.  
 CC -!- SIMILARITY: Contains 15 C2H2-type zinc fingers.  
 CC -!- CAUTION: Ref.4 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF352026; AAK30252.1; -.  
 DR EMBL; AL831845; CAD38551.1; -.  
 DR EMBL; AB037817; BAA92634.1; -.  
 DR EMBL; AC005498; AAC32422.1; ALT\_SEQ.  
 DR Genew; HGNC:23226; ZNF471.  
 DR HSSP; P25490; 1UBD.  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR InterPro; IPR007086; Znf\_C2H2\_sub.  
 DR Pfam; PF01352; KRAB; 1.

DR Pfam; PF00096; zf-C2H2; 15.  
 DR PRINTS; PR00048; ZINCFINGER.  
 DR ProDom; PD000003; Znf\_C2H2; 13.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; ZnF\_C2H2; 15.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 15.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 15.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat.  
 FT DOMAIN 14 85 KRAB.  
 FT ZN\_FING 206 228 C2H2-TYPE 1.  
 FT ZN\_FING 234 256 C2H2-TYPE 2.  
 FT ZN\_FING 262 284 C2H2-TYPE 3.  
 FT ZN\_FING 290 312 C2H2-TYPE 4.  
 FT ZN\_FING 318 340 C2H2-TYPE 5.  
 FT ZN\_FING 346 369 C2H2-TYPE 6.  
 FT ZN\_FING 375 397 C2H2-TYPE 7.  
 FT ZN\_FING 403 425 C2H2-TYPE 8.  
 FT ZN\_FING 431 453 C2H2-TYPE 9.  
 FT ZN\_FING 459 481 C2H2-TYPE 10.  
 FT ZN\_FING 487 509 C2H2-TYPE 11.  
 FT ZN\_FING 515 537 C2H2-TYPE 12.  
 FT ZN\_FING 543 565 C2H2-TYPE 13.  
 FT ZN\_FING 571 593 C2H2-TYPE 14.  
 FT ZN\_FING 599 621 C2H2-TYPE 15.  
 FT CONFLICT 61 61 Y -> C (IN REF. 2).  
 FT CONFLICT 192 192 M -> I (IN REF. 3).  
 FT CONFLICT 406 406 G -> D (IN REF. 3).  
 FT CONFLICT 407 407 V -> A (IN REF. 2).  
 FT CONFLICT 556 556 S -> C (IN REF. 3).  
 SQ SEQUENCE 626 AA; 73009 MW; 7F47ACFB04CE99AA CRC64;

Query Match 92.5%; Score 37; DB 1; Length 626;  
 Best Local Similarity 83.3%; Pred. No. 18;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 ||||:|  
 Db 403 YKGV 408

# RESULT 6

Z319\_MOUSE

ID Z319\_MOUSE STANDARD; PRT; 581 AA.  
 AC Q9ERR8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Zinc finger protein 319.  
 GN ZNF319 OR ZFP319.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RX MEDLINE=21100867; PubMed=11161788;  
 RA Laub F., Aldabe R., Ou J., Ramirez F.;  
 RT "Overexpression of novel zinc-finger protein induces apoptosis in  
 RT NIH3T3 fibroblasts.";  
 RL Genomics 70:375-380(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: May function as a transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF288403; AAG28743.1; -.  
 DR EMBL; BC059823; AAH59823.1; -.  
 DR HSSP; P08045; 1ZNF.  
 DR MGD; MGI:1890618; Zfp319.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 12.  
 DR ProDom; PD000003; Znf\_C2H2; 4.  
 DR SMART; SM00355; ZnF\_C2H2; 12.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 15.  
 KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;  
 KW DNA-binding; Repeat.  
 FT ZN\_FING 75 99 C2H2-TYPE.  
 FT ZN\_FING 103 125 C2H2-TYPE (DEGENERATE).

FT	ZN_FING	131	153	C2H2-TYPE.
FT	ZN_FING	201	223	C2H2-TYPE.
FT	ZN_FING	229	251	C2H2-TYPE.
FT	ZN_FING	257	279	C2H2-TYPE.
FT	ZN_FING	286	308	C2H2-TYPE (DEGENERATE).
FT	ZN_FING	314	336	C2H2-TYPE.
FT	ZN_FING	342	364	C2H2-TYPE.
FT	ZN_FING	370	392	C2H2-TYPE.
FT	ZN_FING	398	420	C2H2-TYPE (DEGENERATE).
FT	ZN_FING	427	449	C2H2-TYPE.
FT	ZN_FING	457	479	C2H2-TYPE (DEGENERATE).
FT	ZN_FING	485	507	C2H2-TYPE.
FT	ZN_FING	513	535	C2H2-TYPE.
FT	ZN_FING	541	563	C2H2-TYPE.
SQ	SEQUENCE	581 AA;	65644 MW;	C78724098648C84F CRC64;

Query Match 90.0%; Score 36; DB 1; Length 581;  
 Best Local Similarity 83.3%; Pred. No. 25;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 :|||||  
 Db 370 FKCGLC 375

# RESULT 7

Z319\_HUMAN

ID Z319\_HUMAN STANDARD; PRT; 582 AA.  
 AC Q9P2F9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 319.  
 GN ZNF319 OR KIAA1388.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI.  
 RT The complete sequences of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:65-73(2000).  
 CC -!- FUNCTION: May function as a transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB037809; BAA92626.1; ALT\_INIT.  
 DR Genew; HGNC:13644; ZNF319.  
 DR HSSP; P08045; 1ZNF.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 12.  
 DR ProDom; PD000003; Znf\_C2H2; 4.  
 DR SMART; SM00355; Znf\_C2H2; 14.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 15.  
 KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;  
 KW DNA-binding; Repeat.  
 FT DOMAIN 8 76 PRO-RICH.  
 FT ZN\_FING 76 100 C2H2-TYPE.  
 FT ZN\_FING 104 126 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 132 154 C2H2-TYPE.  
 FT ZN\_FING 202 224 C2H2-TYPE.  
 FT ZN\_FING 230 252 C2H2-TYPE.  
 FT ZN\_FING 258 280 C2H2-TYPE.  
 FT ZN\_FING 287 309 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 315 337 C2H2-TYPE.  
 FT ZN\_FING 343 365 C2H2-TYPE.  
 FT ZN\_FING 371 393 C2H2-TYPE.  
 FT ZN\_FING 399 421 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 428 450 C2H2-TYPE.  
 FT ZN\_FING 458 480 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 486 508 C2H2-TYPE.  
 FT ZN\_FING 514 536 C2H2-TYPE.  
 FT ZN\_FING 542 564 C2H2-TYPE.  
 SQ SEQUENCE 582 AA; 65546 MW; E2FBD02256BE0A67 CRC64;

Query Match 90.0%; Score 36; DB 1; Length 582;  
 Best Local Similarity 83.3%; Pred. No. 25;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 :|||||  
 Db 371 FKGLC 376

RESULT 8  
 ZN45\_HUMAN  
 ID ZN45\_HUMAN STANDARD; PRT; 682 AA.  
 AC Q02386; P78472;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 45 (BRC1744).  
 GN ZNF45.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=97220056; PubMed=9067431;  
 RA Constantinou-Deltas C.D., Bashiardes E., Patsalis P.C.,  
 RA Hadjimarcou M., Kroisel P.M., Ioannou P.A., Roses A.D., Lee J.E.;  
 RT "Complete coding sequence, exon/intron arrangement and chromosome  
 RT location of ZNF45, a KRAB-domain-containing gene.";  
 RL Cytogenet. Cell Genet. 75:230-233(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 1-400 FROM N.A.  
 RX MEDLINE=92217982; PubMed=1559709;  
 RA Constantinou-Deltas C.D., Gilbert J., Bartlett R.J., Herbstreith M.,  
 RA Roses A.D., Lee J.E.;  
 RT "The identification and characterization of KRAB-domain-containing  
 RT zinc finger proteins.";  
 RL Genomics 12:581-589(1992).  
 CC -!- FUNCTION: May be involved in transcriptional regulation.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -!- SIMILARITY: Contains 1 KRAB domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L75847; AAB05653.1; -.  
 DR EMBL; BC037575; AAH37575.1; -.  
 DR EMBL; M67509; AAA36133.1; -.  
 DR EMBL; M67509; AAA36134.1; -.  
 DR HSSP; P08046; 1A1I.

DR TRANSFAC; T04988; -.  
 DR Genew; HGNC:13111; ZNF45.  
 DR MIM; 601781; -.  
 DR GO; GO:0003700; F:transcription factor activity; TAS.  
 DR GO; GO:0007275; P:development; TAS.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF01352; KRAB; 1.  
 DR Pfam; PF00096; zf-C2H2; 15.  
 DR ProDom; PD000003; Znf\_C2H2; 13.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; Znf\_C2H2; 15.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 15.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 18.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat; Polymorphism.  
 FT DOMAIN 8 78 KRAB.  
 FT ZN\_FING 192 214 C2H2-TYPE.  
 FT ZN\_FING 276 298 C2H2-TYPE.  
 FT ZN\_FING 304 326 C2H2-TYPE.  
 FT ZN\_FING 332 354 C2H2-TYPE.  
 FT ZN\_FING 360 382 C2H2-TYPE.  
 FT ZN\_FING 388 410 C2H2-TYPE.  
 FT ZN\_FING 416 438 C2H2-TYPE.  
 FT ZN\_FING 444 466 C2H2-TYPE.  
 FT ZN\_FING 472 494 C2H2-TYPE.  
 FT ZN\_FING 500 522 C2H2-TYPE.  
 FT ZN\_FING 528 550 C2H2-TYPE.  
 FT ZN\_FING 556 578 C2H2-TYPE.  
 FT ZN\_FING 584 606 C2H2-TYPE.  
 FT ZN\_FING 612 634 C2H2-TYPE.  
 FT ZN\_FING 640 662 C2H2-TYPE.  
 FT VARIANT 187 187 A -> T (in dbSNP:1047452).  
 FT /FTId=VAR\_012019.  
 FT VARIANT 255 255 R -> K (in dbSNP:399098).  
 FT /FTId=VAR\_012020.  
 FT VARIANT 299 299 T -> A (in dbSNP:388706).  
 FT /FTId=VAR\_012021.  
 FT VARIANT 303 303 P -> R (in dbSNP:388685).  
 FT /FTId=VAR\_012022.  
 FT VARIANT 504 504 R -> K (in dbSNP:407731).  
 FT /FTId=VAR\_012023.  
 SQ SEQUENCE 682 AA; 78242 MW; 9C8E536EB00FCE8B CRC64;

Query Match 87.5%; Score 35; DB 1; Length 682;

Best Local Similarity 83.3%; Pred. No. 42;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 |||||  
 Db 472 YKGTTC 477

RESULT 9  
 Z219\_HUMAN

ID Z219 HUMAN STANDARD; PRT; 722 AA.  
 AC Q9P2Y4; Q9BW28;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 219.  
 GN ZNF219.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Testis;  
 RX MEDLINE=20277481; PubMed=10819330;  
 RA Sakai T., Toyoda A., Hashimoto K., Maeda H.;  
 RT "Isolation and characterization of a novel zinc finger gene, ZNF219,  
 RT and mapping to the human chromosome 14q11 region.";  
 RL DNA Res. 7:137-141(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: May function as a transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Ubiquitous.  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AB015427; BAA90526.1; -.  
 DR EMBL; BC000694; AAH00694.1; -.  
 DR Genew; HGNC:13011; ZNF219.  
 DR MIM; 605036; -.  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0003700; F:transcription factor activity; TAS.  
 DR GO; GO:0006351; P:transcription, DNA-dependent; TAS.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 8.  
 DR SMART; SM00355; ZnF\_C2H2; 9.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 6.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat.  
 FT ZN\_FING 57 79 C2H2-TYPE.  
 FT ZN\_FING 85 107 C2H2-TYPE.  
 FT ZN\_FING 163 186 C2H2-TYPE.  
 FT ZN\_FING 274 296 C2H2-TYPE.  
 FT ZN\_FING 302 324 C2H2-TYPE.  
 FT ZN\_FING 498 520 C2H2-TYPE.  
 FT CONFLICT 232 233 MISSING (IN REF. 2).  
 FT CONFLICT 436 436 Q -> E (IN REF. 2).  
 SQ SEQUENCE 722 AA; 76876 MW; B19DA77B148BC45B CRC64;

Query Match 87.5%; Score 35; DB 1; Length 722;  
 Best Local Similarity 83.3%; Pred. No. 44;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKCGLC 6  
 :|||||  
 Db 189 WKCGLC 194

# RESULT 10

## YAA4\_SCHPO

ID YAA4\_SCHPO STANDARD; PRT; 1115 AA.  
 AC Q09798;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C22G7.04 in chromosome I.  
 GN SPAC22G7.04.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbino-witsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*."  
 RL Nature 415:871-880(2002).

CC -!- SIMILARITY: TO YEAST PAN2 AND TO C.ELEGANS F31E3.4.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z54328; CAA91128.1; -.  
 DR PIR; T11614; T11614.  
 DR GeneDB\_SPombe; SPAC22G7.04; -.  
 DR GO; GO:0006378; P:mRNA polyadenylation; ISS.  
 DR GO; GO:0009117; P:nucleotide metabolism; ISS.  
 DR GO; GO:0006464; P:protein modification; ISS.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; ISS.  
 DR InterPro; IPR006055; Exonuclease.  
 DR InterPro; IPR001394; Peptidase\_C19.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00929; Exonuclease; 1.  
 DR Pfam; PF00443; UCH; 1.  
 DR SMART; SM00479; EXOIII; 1.  
 DR SMART; SM00320; WD40; 2.  
 DR PROSITE; PS50235; UCH\_2\_3; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1115 AA; 126919 MW; EEF6F15E1897269D CRC64;

Query Match 87.5%; Score 35; DB 1; Length 1115;  
 Best Local Similarity 66.7%; Pred. No. 65;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 |:::|  
 Db 640 YRCGIC 645



RESULT 11  
 OVO1\_HUMAN  
 ID OVO1\_HUMAN STANDARD; PRT; 235 AA.  
 AC 014753;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative transcription factor Ovo-like 1 (hOvo1) (Fragment).  
 GN OVOL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98051941; PubMed=9383297;  
 RA Chidambaram A., Allikmets R., Chandrasekarappa S., Guru S.C., Modi W.,  
 RA Gerrard B., Dean M.;  
 RT "Characterization of a human homolog (OVOL1) of the Drosophila ovo  
 RT gene, which maps to chromosome 11q13.";  
 RL Mamm. Genome 8:950-951(1997).  
 CC -!- FUNCTION: Putative transcription factor. Involved in hair  
 CC formation and spermatogenesis. May function in the differentiation  
 CC and/or maintenance of the urogenital system (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed in fetal kidney, and also in adult  
 CC pancreas and placenta. Not expressed in intestine, peripheral  
 CC blood lymphocytes and ovary.  
 CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.  
 CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE  
 CC CORRECTED TO EXTEND THE SEQUENCE IN THE N-TERMINAL SO AS TO  
 CC MAXIMIZE THE SIMILARITY WITH THE MOUSE ORTHOLOG. BUT IT WAS NOT  
 CC POSSIBLE TO RECOVER THE FULL SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF016045; AAB72084.1; ALT\_FRAME.  
 DR HSSP; P15822; 1BBO.  
 DR Genew; HGNC:8525; OVOL1.  
 DR MIM; 602313; -.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0003700; F:transcription factor activity; NAS.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
 DR GO; GO:0007283; P:spermatogenesis; NAS.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 4.  
 DR ProDom; PD000003; Znf\_C2H2; 1.  
 DR SMART; SM00355; Znf\_C2H2; 4.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.  
 KW Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat;

KW Transcription regulation.  
 FT NON\_TER 1 1  
 FT ZN\_FING 86 108 C2H2-TYPE 1.  
 FT ZN\_FING 114 136 C2H2-TYPE 2.  
 FT ZN\_FING 142 165 C2H2-TYPE 3.  
 FT ZN\_FING 181 203 C2H2-TYPE 4.  
 SQ SEQUENCE 235 AA; 26388 MW; 511C8A5A7CDDE6FE CRC64;

Query Match 85.0%; Score 34; DB 1; Length 235;  
 Best Local Similarity 83.3%; Pred. No. 25;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YKGLC 6  
 ||| ||  
 Db 142 YKSLC 147

# RESULT 12

## OVO1\_MOUSE

ID OVO1\_MOUSE STANDARD; PRT; 267 AA.  
 AC Q9WTJ2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative transcription factor Ovo-like 1 (mOvol) (mOvola).  
 GN OVOL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Schonbaum C.P., Fantes J., Mahowald A.P.;  
 RT "Characterization of mouse and Caenorhabditis elegans genes related to  
 RT the Drosophila melanogaster ovo/svb gene."  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP FUNCTION.  
 RX MEDLINE=99026118; PubMed=9808631;  
 RA Dai X., Schonbaum C., Degenstein L., Bai W., Mahowald A., Fuchs E.;  
 RT "The ovo gene required for cuticle formation and oogenesis in flies is  
 RT involved in hair formation and spermatogenesis in mice."  
 RL Genes Dev. 12:3452-3463(1998).  
 CC -!- FUNCTION: Putative transcription factor. Involved in hair  
 CC formation and spermatogenesis. May function in the differentiation  
 CC and/or maintenance of the urogenital system.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SKIN, TESTIS, KIDNEY AND WEAKLY  
 CC IN LUNG. NOT DETECTED IN HEART, BRAIN, SPLEEN, LIVER AND SKELETAL  
 CC MUSCLE.  
 CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT E14.5 DAY IN THE  
 CC SUPRABASAL LAYERS OF DEVELOPING EPIDERMIS, AT E15.5 EXPRESSION  
 CC BEGINS IN THE INNER CELLS OF DEVELOPING HAIR GERMS AND RESTRICTED  
 CC TO INNER ROOT SHEATH AND/OR PRECORTICAL CELLS OF DEVELOPING HAIR  
 CC FOLLICLES.  
 CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF134804; AAD29689.1; -.  
 DR EMBL; AF134805; AAD29690.1; -.  
 DR HSSP; P07248; 2ADR.  
 DR MGD; MGI:1330290; Ov011.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0008544; P:epidermal differentiation; IMP.  
 DR GO; GO:0007498; P:mesoderm development; IMP.  
 DR GO; GO:0007283; P:spermatogenesis; IMP.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 4.  
 DR ProDom; PD000003; Znf\_C2H2; 1.  
 DR SMART; SM00355; Znf\_C2H2; 4.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 3.  
 KW Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat;  
 KW Transcription regulation.  
 FT ZN\_FING 118 140 C2H2-TYPE 1.  
 FT ZN\_FING 146 168 C2H2-TYPE 2.  
 FT ZN\_FING 174 197 C2H2-TYPE 3.  
 FT ZN\_FING 213 236 C2H2-TYPE 4.  
 SQ SEQUENCE 267 AA; 30221 MW; DAD4F51150C21C2D CRC64;  
  
 Query Match 85.0%; Score 34; DB 1; Length 267;  
 Best Local Similarity 83.3%; Pred. No. 28;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 ||| ||  
 Db 174 YKSLC 179

RESULT 13  
 Z185\_MOUSE  
 ID Z185\_MOUSE STANDARD; PRT; 352 AA.  
 AC Q62394;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Zinc finger protein 185 (LIM-domain protein ZFP185) (P1-A).  
 GN ZNF185 OR ZFP185.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96425694; PubMed=8828036;  
 RA Levin M.L., Chatterjee A., Pragliola A., Worley K.C., Wehnert M.,

RA Zhuchenko O., Smith R.F., Lee C.C., Herman G.E.;  
 RT "A comparative transcription map of the murine bare patches (Bpa) and  
 RT striated (Str) critical regions and human Xq28.";  
 RL Genome Res. 6:465-477(1996).  
 CC -!- FUNCTION: May be involved in the regulation of cellular  
 CC proliferation and/or differentiation (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- TISSUE SPECIFICITY: Expressed in skin, kidney, ovary, testis. Also  
 CC expressed in brain, cartilage, heart, lung, spleen and thymus.  
 CC -!- DEVELOPMENTAL STAGE: At E14.5, only expressed in mesenchymal  
 CC cells. At E16.5 expressed also in cells lining the vertebrae and  
 CC tendons of the proximal tail. In late embryogenesis, expressed in  
 CC mesenchymal cells adjacent to the distal limb bones (tibia and  
 CC calcaneum), in tendons and in the connective tissue sheath  
 CC (epimysium) surrounding the skeletal muscle. Also expressed in the  
 CC epithelia of the epididymis of the testis.  
 CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U46687; AAC52628.1; -.  
 DR MGD; MGI:108095; Zfp185.  
 DR InterPro; IPR001781; LIM.  
 DR ProDom; PD000094; LIM; 1.  
 DR SMART; SM00132; LIM; 1.  
 DR PROSITE; PS00478; LIM\_DOMAIN\_1; FALSE\_NEG.  
 DR PROSITE; PS50023; LIM\_DOMAIN\_2; 1.  
 KW LIM domain; Metal-binding; Zinc.  
 FT DOMAIN 23 26 POLY-GLU.  
 FT DOMAIN 292 347 LIM.  
 SQ SEQUENCE 352 AA; 38322 MW; 2AB1F833D7AF1A5C CRC64;

Query Match 85.0%; Score 34; DB 1; Length 352;  
 Best Local Similarity 66.7%; Pred. No. 35;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YKCGLC 6  
 :|||:|  
 Db 319 FKCGIC 324

#### RESULT 14

##### Z185\_HUMAN

ID Z185\_HUMAN STANDARD; PRT; 457 AA.  
 AC O15231; O00345; Q9NSD2;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Zinc finger protein 185 (LIM-domain protein ZNF185) (P1-A).  
 GN ZNF185.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=97422610; PubMed=9268636;  
 RA Heiss N.S., Gloeckner G., Baechner D., Kioschis P., Klauck S.M.,  
 RA Hinzmann B., Rosenthal A., Herman G.E., Poustka A.;  
 RT "Genomic structure of a novel LIM domain gene (ZNF185) in Xq28 and  
 RT comparisons with the orthologous murine transcript."  
 RL Genomics 43:329-338(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20314869; PubMed=10854409;  
 RA Mallon A.M., Platzer M., Bates R., Gloeckner G., Botcherby M.,  
 RA Nordsiek G., Strivens M.A., Kioschis P., Dangel A., Cunningham D.,  
 RA Straw R., Weston P., Hunter C., Gilbert M., Fernando S., Goodall K.,  
 RA Kerry G., Greyststrong J.S., Clark D., Goerdes M., Blechschmidt K.,  
 RA Rump A., Hinzmann B., Mundy C.R., Miller W., Poustka A., Herman G.E.,  
 RA Rhodes M., Denny P., Rosenthal A., Brown S.D.M.;  
 RT "Comparative genome sequence analysis of the Bpa/Str region in mouse  
 RT and man."  
 RL Genome Res. 10:758-775(2000).  
 CC -!- FUNCTION: May be involved in the regulation of cellular  
 CC proliferation and/or differentiation.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- TISSUE SPECIFICITY: Expressed in placenta, pancreas and kidney.  
 CC Also expressed in prostate, testis, ovary and blood.  
 CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y09538; CAA70733.1; -.  
 DR EMBL; U82671; -; NOT\_ANNOTATED\_CDS.  
 DR Genew; HGNC:12976; ZNF185.  
 DR MIM; 300381; -.  
 DR GO; GO:0008270; F:zinc ion binding; TAS.  
 DR InterPro; IPR001781; LIM.  
 DR ProDom; PD000094; LIM; 1.  
 DR SMART; SM00132; LIM; 1.  
 DR PROSITE; PS00478; LIM\_DOMAIN\_1; FALSE\_NEG.  
 DR PROSITE; PS50023; LIM\_DOMAIN\_2; 1.  
 KW LIM domain; Metal-binding; Zinc.  
 FT DOMAIN 24 30 POLY-GLU.  
 FT DOMAIN 397 452 LIM.  
 FT CONFLICT 1 9 MTTEDYKKL -> MQRQ (IN REF. 1).  
 FT CONFLICT 176 176 Y -> C (IN REF. 1).  
 FT CONFLICT 182 182 A -> P (IN REF. 1).  
 FT CONFLICT 260 260 G -> R (IN REF. 1).  
 SQ SEQUENCE 457 AA; 49187 MW; 1E8A77D6B75EB6C6 CRC64;

Query Match 85.0%; Score 34; DB 1; Length 457;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKCGLC 6  
:||||  
Db 424 FKCGIC 429

RESULT 15

LI26\_CAEEL

ID LI26\_CAEEL STANDARD; PRT; 472 AA.  
AC Q27355;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transcription factor lin-26.  
GN LIN-26 OR F18A1.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=95044920; PubMed=7956818;  
RA Labouesse M., Sookhareea S., Horvitz H.R.;  
RT "The Caenorhabditis elegans gene lin-26 is required to specify the  
RT fates of hypodermal cells and encodes a presumptive zinc-finger  
RT transcription factor."  
RL Development 120:2359-2368(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Kirsten J.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Presumed to be a transcription factor required to  
CC specify the fates of hypodermal and neuron-associated support  
CC cells.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z32673; CAA83585.1; -.  
DR EMBL; U41535; AAB63406.1; -.  
DR PIR; T34222; T34222.  
DR TRANSFAC; T03602; -.  
DR WormPep; F18A1.2; CE04402.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 1.

DR SMART; SM00355; ZnF\_C2H2; 1.  
 KW Transcription regulation; Zinc-finger; Repeat; Metal-binding;  
 KW DNA-binding; Nuclear protein.  
 FT ZN\_FING 335 363 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 369 398 C2H2-TYPE (ATYPICAL).  
 FT DOMAIN 399 465 ASP/GLU-RICH (ACIDIC).  
 FT MUTAGEN 350 350 G->E: IN LIN-26(MC1); EMBRYONIC LETHALITY  
 FT AND THE DEATH OF MANY HYPODERMAL CELLS.  
 FT MUTAGEN 351 351 S->F: IN LIN-26(MC2); EMBRYONIC LETHALITY  
 FT DUE TO A FAILURE OF HYPODERMAL CELLS TO  
 FT ENCLOSE THE EMBRYO.  
 FT MUTAGEN 355 355 L->F: IN LIN-26(N156); HYPODERMAL PN.P  
 FT CELLS EXPRESS A NEURAL FATE; ANIMALS LACK  
 FT A VULVA.  
 SQ SEQUENCE 472 AA; 53203 MW; EE1101DDE4AF0AE7 CRC64;

Query Match 85.0%; Score 34; DB 1; Length 472;  
 Best Local Similarity 83.3%; Pred. No. 46;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YKGLC 6  
 ||| ||  
 Db 335 YKCALC 340

Search completed: March 5, 2004, 16:23:40  
 Job time : 1.81481 secs